5-2010

THE ATP WORLD TOUR: HOW DO PRIZE STRUCTURE AND GAME FORMAT AFFECT THE OUTCOME OF A MATCH?

Carolina Salge
Clemson University, csalge@clemson.edu

THE ATP WORLD TOUR:
HOW DO PRIZE STRUCTURE AND GAME FORMAT AFFECT THE OUTCOME OF A MATCH?

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Masters of Arts
Economics

---

by
Carolina Salge
May 2010

Accepted by:
Dr. John T. Warner, Committee Chair
Dr. Michael T. Maloney
Dr. Robert D. Tollison

ABSTRACT

Income differences are normally explained by theory in terms of marginal productivity. However, how can one explain salary enormous jump when a vice-president is promoted into a CEO position? It would be too naive for one to believe his/her productivity increased at the same rate as his/her salary. Differences in wage in terms of relative performance is what economists call tournament theory. What we normally see is that workers are underpaid when they enter a company, and then become overpaid after a promotion. The main point behind such theory is that beginners work hard in order to enjoy the fat paycheck later on. Hence, the large increase in a CEO's salary covers all the effort he/she supplied when underpaid.

Tournament theory has been tested and applied in many professional sports. This thesis will analyze the data gathered from the Association of Tennis Professionals website. Its focus will be on the outcomes of a tennis match. Six main hypotheses will be tested: (1) expected marginal payoffs is expected to increase at a linear or at an increasing rate up to the final round, and then have a distinct jump in the final match; (2) best 4 players of a tournament (champion, finalist, and 2 semi-finalist) are expected to receive 50% of total prize money; (3) the probability a player wins a match is expected to increase if he wins the first set. The effect of winning the first set is expected to be more important for lower ranked players; (4) upsets are more likely to occur in ATP tournaments (played in best of 3 sets) when compared to Grand Slam tournaments (played in best of 5 sets); (5) a change in spread level is expected to directly influence the outcome of a match. As spread level increases more wins coming from lower skilled

players is expected; (6) total prize money will not affect the outcome of a match but will change the entry in a tournament. Better players are expected to enter tournaments where total purse is higher; (7) The effect of outcome is expected to depend on players' ranking level. Less competition is expected to be seen when ranking difference gets wider.

The construction of marginal payoff in professional tennis tournaments does not follow the theory. Marginal payoffs do increase, but at a decreasing rate, and an even larger decrease in the final round. Also, marginal payoff became negative, sometimes. In most cases, top four players do not receive 50% of total purse, but around 43%.

The probability a player wins a match does increase after he wins the first set. The effect of winning the first set was the same in percentage terms for both better and worse players; yet, in terms of ratio difference, first set is doubled important for worse players. Results from binomial distribution showed that upsets are more likely to occur in matches played in best of 3 sets when compared to matches played in best of 5 sets. An increase in amount of set played does give an advantage for the better player.

Output from regressions on outcome show that a change in spread level will increase the number of upsets in a tournament, meaning that payoff seems to be more important for worse ranked players. Regression results on average rank weight pointed out that an increase in tournament total purse will increase the quality of the draw, which is consistent with theory. In terms of ranking, variables Top Ranking and Bottom Ranking explain results in outcome in a more efficient way, rather than variable Ranking Difference. Lastly, results lead to the conclusion that less competition is seen as differences in ranking increase. Dominance does exist for the better player ranked in the top 10 when compared to the rest subgroups of better ranked players.

DEDICATION

I primarily dedicate this thesis to my parents, Adriano and Luzia Maria, my brother, Cassio, and my boyfriend, Thomas. I also dedicate this thesis to my friends Michael McLeod, Nancy Harris, Jose Caban, Daniela Alvarez, Maria Brito, and Alexandra Luc. Thank you for everything you ever done for me. Still, I want to dedicate this thesis and thank my uncle Tarquinio Lucio and my aunt Ana Claudia for coming all the way from Brazil for both of my graduations (bachelor and masters). Lastly, I want to dedicate this thesis to everyone that contributed a little bit of their time to help me get where I am today. Thank you!

ACNOWLEDGMENTS

Firstly, I want to thank Professor Warner for his teaching excellence that had a significant contribution to my education in economics. I am also very grateful to Professor Warner for his endless attention and effort during my master thesis process at Clemson University. Finally, I would like to thank the members of my committee, Professors Michael Maloney, and Robert Tollison for their precious input regarding my thesis.

TABLE OF CONTENTS

LIST OF TABLES

List of Tables (Continued)

List of Tables (Continued)

LIST OF FIGURES

INTRODUCTION

Literature on tournament theory is extremely robust with sports-related studies.[1] Most recent analyses' focus was on testing the tournament model developed by Rosen (1986), where players adjust effort level in order to maximize their expected prize or expected utility.[2] Rosen's theory predicts that individuals are expected to work harder if spread becomes wider. However, can we say that Rosen's theory is applied in the real world of professional men's tennis?

The amount of total prize money in professional tournaments is fixed, and determined ahead of time. Hence, players are aware of spread levels before entering a tournament. If a player's effort level depends on the size of spread, as Rosen says, then one expects to see better players entering tournaments where spread is wider. Not only that but an increase in spread level ought to decrease the amount of upsets (when a higher seeded player loses to a lower seeded player (or unseeded) player). Theory predicts that differences in marginal payoff (discounted sum of the differences in spread) need to be increasing at an increasing rate, if players are risk averse, and have an even higher jump in the final match, in order to maintain players interested in winning. Rosen (1986) mentioned that 50% of total prize money is distributed to the top four players in a men's professional tennis tournament, but is that still true? Do we see all of these assumptions in the tour?

---

[1] See Nalebuff and Stiglitz (1983), Green and Stokey (1983), Main, O'Reilly, and Wade (1993), Knoeber and Thurman (1994), Ivankovic (1995), and Lazear (1996)

[2] See Melton and Zorn (2000), Lynch and Zax (2000), Maloney and Terkun (2002), Sunde (2003), Bentley and Maloney (2006), and Gilsdorf, and Sukhatme (2007) just to list a few.

In this thesis, I propose to answer these questions by looking at data on mens' professional tennis tournaments. Perhaps, professional tennis tournaments are the best fit for tournament model since they are structured on a single elimination basis, where winners qualify to the next round and losers are out of competition. This thesis will also analyze descriptive statistics in mens' professional tennis, for example, if there are more upsets in matches played best of 3 sets compared with matches played best of 5 sets, and how does winning probability changes after a player wins the first set.

As Rosen points out, spread levels do exist and need to be correctly structured in order to keep a high level of players' performance. This study is extremely important when maximizing utility level of a player, as tournament directors might want to balance players' marginal benefit of moving into the next round (future prize) against players' marginal cost of moving into the next round (effort level), in order to attract audience. The data I gathered is rich because I know outcome of matches, players' ranking, total prize money of tournaments and their spread levels, surface in which match was played, and number of sets (if played best of 3 or 5).

Figure I: Percentage Average of Total Purse Distributed to Winners

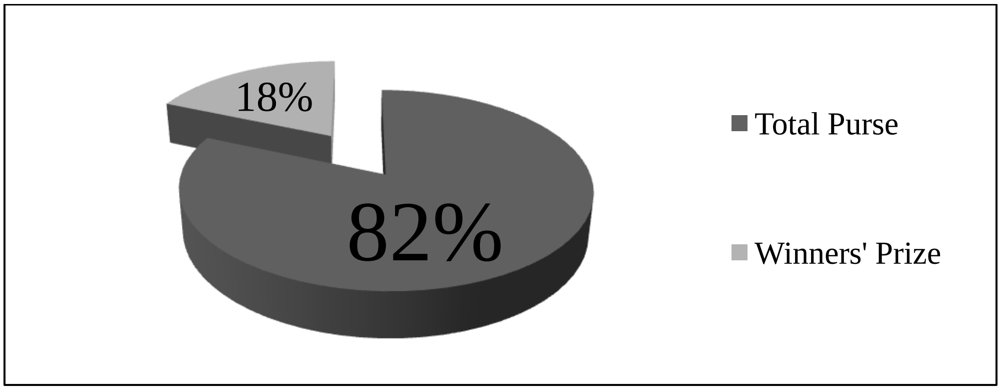


Figure 1, for example, shows the average percentage of total purse distributed to the top players in the ATP World Tour tournaments. Last round pay in mens' professional tennis is similarly structured to the last round pay in professional golf (Senior PGA Tour), where instead of receiving an average of 18% of total purse winners in golf receive 19%.

Part 2 will discuss review of literature. Part 3 provides information about the Association of Tennis Professionals (ATP). Part 4 describes dataset of the ATP World Tour. Part 5 will discuss hypotheses tested in this thesis. Part 6 will show results derived from assumptions in part 5. Part 7 grants the reader with concluding thoughts while parts 8 and 9 present the appendix and bibliography of this study.

REVIEW OF LITERATURE

Tournament Theory

Twenty-four years ago, Rosen (1986) developed a model to explain the problem of structuring incentives when competitors are paid on the basis of rank or relative performance. Rosen investigated the incentive properties of prizes in sequential elimination tournaments, where rewards increase in each stage of play. Rosen mentioned that prize structure needs to promote "survival of the fittest" and "quality of play" in every stage of the competition. His design of the game (professional tennis) assumes that all tournaments start with 2^N players and is followed by N matches. Champions' purse is defined by W_1, finalists' purse is defined by W_2. Purse of semi-finals is W_3, and so on up to the first round. The number of rounds left to be played is defined by s. All players who lost in match s stages where remained will get prize equal to W_{s+1}. Marginal Payoff of advancing a round was determined as $\Delta W_s = W_s - W_{s+1}$. Rosen mentioned that a player's decision of how much effort to supply will depend on the benefit of higher effort against its costs. Let I be defined as ability level supplied by player of type i, and J defined as ability level supplied by player of type j. Both I and J take on m possible values, 1,2,…,m, with m $\leq 2^N$. Let x_{si} and x_{sj} be the intensity of effort expended by players i and j in a match where s stages are left to be played, and let γ_I and γ_J be the natural talents for the game. Then, the probability a player of type I will defeat a player of type J is:

$$P_s(I,J) = \gamma_I h(x_{si}) / \gamma_I h(x_{si}) + \gamma_J h(x_{sj}) \ , \qquad (1)$$

where h(x) is increasing in x and h(0)≥0. A player will increase his winning probability by supplying a higher level of effort, given the talent and effort of the opponent and own talent. When both player supply same level of effort, winning probability will be:

$$P_s(I,J) = \gamma_I / (\gamma_I + \gamma_J) \tag{2}$$

Equation (2) simply says that a player's probability of winning is proportional to his ability relative to the sum of both players' ability. A player's decision of how much effort to supply in a match will depend on its benefit minus its cost. According to Rosen, there are two complications: (1) value of advancing a round depends on how the player assesses future effort; (2) current behavior will depend on future possible opponents. Hence, the fundamental equation for players' strategies is

$$V_s(I,J) = \max[P_s(I,J)EV_{s-1}(I) + (1-P_s(I,J))W_{s+1} - c(x_{si})] \tag{3}$$

where $V_s(I,J)$ is the value to a player of type I playing a match against a player of type J in the next round. The cost of effort supplied by players is c(x) where players choose x_s to maximize V_s. The expected value of eligibility in the next round is $EV_{s-1}(I)$. When observing $V_s(I,J)$ what we see is that players' strategies consist of an equation that equals marginal benefit of advancing to next round $\max P_s(I,J)EV_{s-1}(I) + (1-P_s(I,J))W_{s+1}$ minus marginal cost of supplying effort $c(x_{si})$. Substituting the winning probability equation where both players supply same amount of effort (2) into the equation for players' strategies (3), and differentiating with respect to intensity of effort (x_{si}) yields the first order condition

$$\gamma_I \gamma_J h_j h'_i / (\gamma_I h_i + \gamma_J h_j)^2 [EV_{s-1}(I) - W_{s+1}] - c'_i = 0 \qquad (4)$$

where $h_i = h(x_{si})$, $h'_i = \mathrm{d}\, h(x_{si}) / \mathrm{d}\,(x_{si})$, etc. In equation (4) effort in controlled by the expected value of eligibility in the next round minus the prize earned if the match is lost $(EV_{s-1}(I) - W_{s+1})$. The difference between the expected value of eligibility in the next round must be higher than the prize paid if player loses in current round, in order to keep the player interested into moving to the next round.

The second order condition is

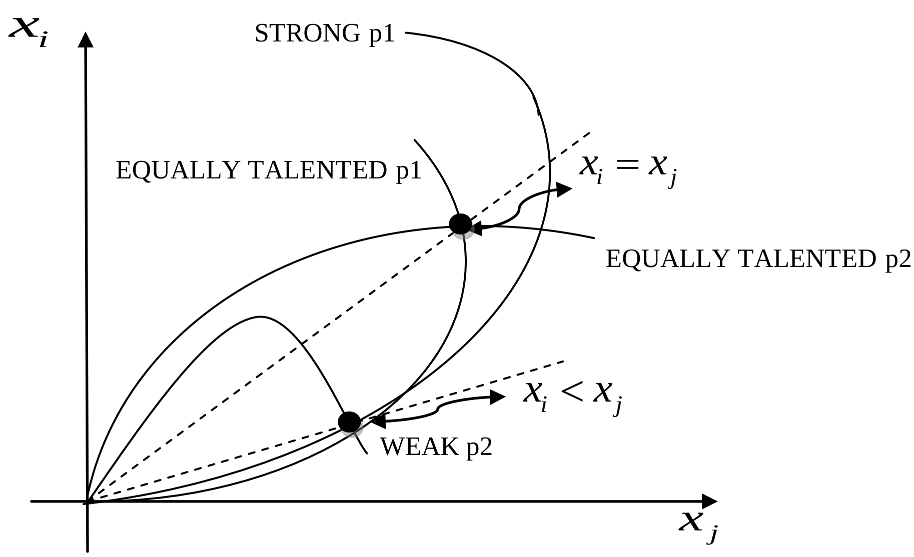
$$D = c'_i [(h''_i / h'_i) - 2\gamma_I h'_i / (\gamma_I h_i + \gamma_J h_j)] - c''_i < 0. \qquad (5)$$

where player's marginal cost of moving to the next round is increasing at a faster rate than his marginal benefits, causing D<0. For example, a weak player might just slack off against a strong one because he knows that even though he supplies his maximum level of effort, that it will not be enough to beat his opponent. Differentiating equation (4) with respect to the current opponent's effort,

$$\partial x_i / \partial x_j = \{(h'_j / h_j) c'_j / -D(\gamma_I h_i + \gamma_J h_j\}(\gamma_I h_i - \gamma_J h_j). \qquad (6)$$

Player i's best response will be to increase effort (x_j) when x_j is small enough, but it will be to decrease effort when opponent's effort is significantly large. In Figure II, when $x_i = x_j$ players 1 and 2 are equally talented, hence, both of them will exert same amount of effort. On the other side, when $x_i < x_j$ player 1 is better than player 2, which will cause player 2 to exert a higher level of effort at the beginning of the match, but after player 2 reaches his peak and sees he has no chances of winning the match, he will completely decrease amount of effort; player 1 will increase amount of effort from the beginning of the match, at a lower rate when compared to player 2, and after player 1 realizes he is the better player he will keep increasing effort at a decreasing rate. Note that player 1 will never reach his peak, meaning that he will play just well enough to win the match.

Figure II: Pure Strategy Equilibra



Elasticities from (6) can be defined as

$$\begin{aligned} \eta(x) &= xh'(x)/h(x), \\ \varepsilon(x) &= xc'(x)/c(x), \\ \mu(x) &= \eta(x)/\varepsilon(x). \end{aligned} \tag{7}$$

where $h(x)$ is a function that translates effort into probability of winning, and $c(x)$ is a function that translates effort into the cost of winning. If the elasticity of response in probability of winning is higher than the elasticity of response in cost, players are expected to supply more effort, even though their utility level might go down. $\mu(x)$ is performance/cost ratio at the level of x. Performance/cost ratio here is defined as the slope of expected payback in effort to the slope of expected cost in effort at the level x.

Applying Model for Equally Talented Players

Turning point occurs when equilibrium is $x_i = x_j$ for equally talented players (P=1/2). If $x_i < x_j$ and $\gamma_I > \gamma_{J,}$ player I will face a weak opponent. However, if $x_i > x_j$ and $\gamma_I < \gamma_{J,}$ then player I will face a strong player. When all players are equally talented $EV_{s-1}(I)$ will equal V_{s-1} because players know they will face opponents of same skills, in all rounds. Assuming winning probability equal to 1/2, yields the effort level equation, which is the marginal benefit equals the marginal cost, when s matches are remained to be played:

$$(V_{s-1} - W_{s+1})(h'(x_s)/h(x_s)/4 = c'(x_s). \tag{8}$$

Substituting elasticities found in (7) into (8):

$$(V_{s-1} - W_{s+1})\mu(x_s)/4 = c(x_s). \tag{9}$$

Substituting (9) into (2) and again using P=1/2:

$$\begin{aligned} V_s &= \{(1/2)(1-\mu(x_s)/2\}(V_{s-1} - W_{s+1}) + W_{s+1} \\ &= \beta_s V_{s-1} + (1-\beta_s)W_{s+1}, \end{aligned} \tag{10}$$

where, $$\beta_s = (1/2)(1-\mu(x_s)/2). \tag{11}$$

Equation (10) is implying that for given prizes, players are better off by supplying more effort, even if their utility level decreases, because if they don’t supply more effort we would have a default. Rules must be balanced to control for actions that influence outcome in an inefficient and unproductive way (competition will be destroyed if rules are not controlled)[3]. However, equation (10) will only hold if players have zero incentive to default from x_s defined by (8). This entails, from (10), that the incentives of moving to the next round must be higher than zero. $V_{s-1} - W_{s+1} = \beta_s (V_{s-1} - W_{s+1}) > 0$, which means that β_s >0, or else, accepting a loss will be the better choice. There is no equilibrium if players have incentives to default a match. Assuming 0< β_s <1 for all s and that probability of advancing round s equal to loser's pay in round s, will define equation (10) as the marginal payoff players receive by advancing a round:

$$V_s = (\beta_1 \beta_2 ... \beta_s)\,\Delta W_1 + (\beta_2 ... \beta_s)\,\Delta W_2 + ... + \beta_s\,\Delta W_s + W_{s+1}\,. \qquad (12)$$

The incentive of keep winning in a tournament comes from marginal payoff, which is the prize a player will get plus the discounted sum in prize difference that might be achieved in future matches. Manipulating values in (12) gives the expression for the probabilities player I will face player J in the next round minus the prize earned if the match is lost $(V_{s-1} - W_{s+1})$,which controls performance incentives, from (9). The results can be written as

$$(V_s - W_{s+1}) = (\beta_1 ... \beta_{s-1})\Delta W_1 + (\beta_2 ... \beta_{s-1})\Delta W_2 + ... + \Delta W_s\,. \qquad (13)$$

Constant performance will require that $(V_s - W_{s+1})$ is a constant, which Rosen calls k. Hence, the following derivation comes from equation (13)

$$k(1-\beta) = \Delta W = \Delta W_s \text{ for s=2, 3,…, N.} \qquad (14)$$

[3] There will be no competition if all players decide to default. Moreover, players have incentives to use innovative techniques to increase their winning probability (steroids for instance), and that's also why organizations' like the ATP create rules to maintain the integrity, and competition of the sport.

Because the final match depends only on ΔW_1, from (14)

$$k = \Delta W_1 = \Delta W/(1-\beta) > \Delta W. \tag{15}$$

From equation (15) Rosen concluded that prize distribution weights top prize money more greatly than the rest. It is clear that concentrating even more in the top purse creates incentives for performance to increase as the game proceeds through its stages (Rosen, 1986). If players are risk averse, difference in the final spread will need to be even larger, in order to keep players interested. Rosen also developed models for heterogeneous contestants with known talents, and heterogeneous contestants with talents unknown. When the scenario involves different players with known abilities Rosen expects to see a lower level of effort supplied by players. In other words, the larger the initial disadvantage of a player, the harder it will be for him, in terms of effort, to compensate his weakness. On the other side, a strong player knows about his initial advantages and is able to reduce effort level without excessively jeopardizing his chances of success. In terms of marginal payoff, a weak player will have an inferior level of marginal payoff, because his probability of winning a match is lower, when compared to a strong player. However, winning probabilities are expected to average out as the number of rounds remained to be played in a tournament decrease.

When ability level of players is unknown it is in the interest of the better player to make his rivals believe his abilities are higher than it normally is, in order to induce a rival to supply less effort (he doesn't believe he has a chance to win). For weak players, it is smart to make their rivals believe he is weaker than he normally is, in order to induce the rival to slack off**.** Marginal payoff for players with unknown abilities will be the same prior the start of a match because one cannot predict a winning probability as all players will choose same strategy.

Empirical Studies on Tournament Theory

Many economists have conducted empirical studies to test Rosen's work in different types of labor markets, with a particular focus on professional sports. Eriksson (1999) tested the predictions of tournament theory using data set on Danish executives. His results pointed to the acceptation of tournament theory as he found that managers having important duties inside the firm did have a wider salary spread. DeVaro (2005) suggested that relative work performances determined promotions. Higher work performance will increase the chance of a worker getting promoted. He then estimated a structural model of promotion tournaments (based on worker performance, wage spread between promotions, and promotion as an endogenous variable) and accepted the predictions of tournament theory, where employers are motivated by larger spreads.

Sunde (2003) conducted an empirical study on the real financial incentive on professional tennis players based on effort level, using data from last two rounds of Grand Slam and Masters Series tournaments. He found a significant positive relationship among spread level and amount of games won by a player in a match, and no financial incentive in terms of total purse. Melton and Zorn (2000) analyzed the influence of tournament participation on players' incentive level on the Senior PGA Tour. The format in professional golf (Senior PGA Tour) ensures that all players signed in the tournament, where they value the cut, will receive a portion of total purse, regardless of their ranking.

Their results support the idea that larger prize level of tournaments improves players' performance and incentives.

Coffey and Maloney (2006) compared tournament theory with pure selection model in horse and dog racing. Their findings show that not only the best fit was tournament theory, but that nearly two-thirds of the increased performance, based on effort level, was associated with higher prizes. Maloney and Terkun (2002) investigate prize structures among competing firms paying tournament wages in motorcycle racing. Their study argued that even though some researchers found that workers supply more effort when there is a wider spread, firms seem to recognize this. Hence, firms respond to workers' reaction by offering higher expected wages through wider purses.

Ivankovic (1995) answered three major issues in tennis tournaments: (1)does the level of marginal payoff matter regarding players' amount of effort, (2)should the level of total purse determine who enters a tournament, and (3)is is true that players' effort only depend on the marginal payoff, and not total purse. He found that marginal payoff was not structured as theory predicts. Marginal Payoff level increased at a decreasing rate and in the final round it became negative. However, he observed that as spread level was increased, player supplied a higher level of effort (time length of matches increased). On the other side, Ivankovic was able to accept the hypothesis that total purse influences entry in a tournament. As the level of purse increases, better players are more likely to entry tournament. There was no confidence in saying that purse level does not affect effort supplied by players as he got mixed results.

The Binominal Distribution

The binomial distribution is the discrete probability distribution of the number of successes in a sequence of *n* opportunities with probability of success equals to *p*. The discussion about the binomial distribution is useful to this thesis because in hypothesis 4 I will be testing if upsets are more likely to occur in ATP tournaments (played in best of 3 sets) when compared to Grand Slam tournaments (played in best of 5 sets). Applying the binomial distribution to my hypothesis yields the following formula:

$$p = X / n \text{ = number of sets won/ number of sets in match} \qquad (16)$$

since, the variance of p [$V(p) = P(1-P)/N$] decreases as *n* increases there will be less variance in outcome. In other words, an increase in the number of sets played is expected to increase the probability that the better player will win the match.

Study of Binomial Distribution in Professional Tennis

Dinterman (2009) tested the null hypothesis that there is a difference in the frequency of upsets in matches when played best of 3 sets verses matches played best of 5 sets. His dataset included all matches played in Grand Slam tournaments from 2001 until 2008, and all matches played in seven major ATP tournaments from 2001 until 2008. Upset was defined as higher seeded player losing to a lower seeded player (or unseeded) player. His results showed a higher frequency of upsets in tournaments played in 3 sets than in tournaments played in 5 sets.

ATP

The empirical study of this thesis was based on the data set gathered from the Association of Tennis Professionals' (ATP) website. The Americas headquarters of the ATP is in Ponte Vedra Beach, Florida, while ATP Europe is located in Monaco. ATP International is based in Sydney, and it covers Africa, Asia, and Australia. The ATP Executive Offices are in London. Data located on the ATP website include current and past rankings, prize money, live scores, daily scores, draws, results archive, service game leaders, return of services leaders, first serve points won, second serve points won, service games won, break points saved, head-to-head competition, tournament history, singles and doubles activity by player, and match statistics. All the data used in this study were collected from the ATP website.

The ATP was formed in 1972, and its main purpose is to protect the interest of male professional tennis players. After January 2009 the ATP was renamed, and is now knows as the ATP World Tour. Jack Kramer and Cliff Drysdale first managed the ATP organization. In 1991, the ATP had its first TV Broadcasting package that included 19 tournaments. Their first website was created in 1995, and it was rapidly boosted by a contract with Mercedes-Benz. In 2008, the ATP tour was restructured after facing couple lawsuits.

The ATP Tour embraces six different types of tournaments: Grand Slams, ATP Tour Masters 1000, ATP World Tour 500 Series, ATP World Tour 250 Series, ATP Challenger Series, and Futures Tournaments. In this thesis only four types are being observed and tested: Grand Slams, Masters 1000, 500 and 250 Series. The ATP also runs the World Team Cup, seeded by Düsseldorf every May, and the ATP Champions Tour for seniors. At the end of every season the ATP and the ITF organize the Tennis Masters Cup, which includes only the best eight (singles and doubles) players in the world.

The Grand Slams are the strongest tournaments played on the ATP Tour. Only four of them are played each year (Australian Open, Roland Garros, Wimbledon, and the US Open). The Masters 1000 is the second strongest division of tournaments played on the ATP Tour. There are nine tournaments played every year (Indian Wells, Miami, Monte Carlo, Rome, Madrid, Montreal, Cincinnati, Shanghai, and Paris. The Tour Finals is played in London. The 500 Series is the third highest-level tournaments played on the ATP Tour. There are eleven events of this port and they are played in Rotterdam, Dubai, Acapulco, Memphis, Barcelona, Hamburg, Washington, Beijing, Tokyo, Basel, and Valencia. In 2009, the Davis Cup World Group and the World Group Playoffs began to award a total of 500 points to players. The 250 Series is the lowest tier of tournaments played on the ATP Tour. The series includes forty tournaments that take place in all continents of the world.

The ATP updates both of its rankings, Entry ranking and Race ranking, every week. The entry ranking determines qualification for entry and seeding in all tournaments for both singles, and doubles that accumulates from year to year. ATP Race ranking was an annual race from beginning of a season to the end of a season but was removed in the beginning of 2009.

DATA SET

All the data set were gathered by the author through the Association of Tennis Professionals' website (ATP). The data consists of observations on individual matches in 129 professional tennis tournaments played on the ATP Tour during the years of 2008 and 2009. Tournaments differ in year; draw size based on the number of players participating, and rounds based on the number of matches left to be played. Year is defined as matches played on 2008 or 2009. There are seven different draw sizes 28 draw, 32 draw, 48 draw, 56 draw, 64 draw, 96 draw, and 128 draw. The 96 draw is the typical draw size for the Masters Series tournaments, and the 128 draw is normally the four Grand Slam tournaments. The draw not only lists each player's names and the opponent's name but it also shows possible opponents a player might face each round up to the final match.

There are 18 28-draws, 78 32-draws, 8 48-draws, 15 56-draws, 1 64-draw, 4 96-draws, and 8 128-draws. Overall, there are 5, 280 observations on individual matches in 129 tournaments. The data set is divided in two groups. First group is the Tournament Data Set, which includes all the significant information regarding a tennis match. Second group is the Marginal Payoff and Prize Money Data Set, and it contains tournament financial information.

The Tournament Data Set variables are: TOURNAMENT (a number given for each number, the maximum is 129), TOURNAMENT NAME (name of tournament given by the ATP), YEAR (what year tournament was played), SETS (1 if played best of 5 sets and 0 if played best of 3 sets), ROUND (sequence of matches played, from 1 to 7), DRAW (the draw format the tournament follows, it can either be 28, 32, 48, 56, 96, or 128), TOP RANKING (top ranked player's ranking when the match was played, according to the ATP), BOTTOM RANKING (bottom ranked player's ranking when the match was played, according to the ATP), TOP RANKED PLAYER (top ranked player's name), BOTTOM RANKED PLAYER (bottom ranked player's name), FIRST SET WINNER (1 if top player won the first set, 0 if bottom player won the first set, and 3 if first set was incomplete), OUTCOME (1 if top player won the match and 0 if bottom player won the match), SURFACE (1 if played on hard court, 2 if played on clay court, 3 if played on grass court, and 4 if played on carpet court), RANK DIFF (absolute value of difference between two players ranking), RANK WEIGHT (sum of all players ranking in a determined tournament), and AVERAGE RANK WEIGHT (division of rank weight by draw size).

Quite a few variables were created for empirical purposes. Variable RANK DIFF was created in order to measure top player's ability with respect to bottom player's ability. RANK WEIGHT was generated in order to measure players level in a tournament; the higher the number, the lower the level. AVERAGE RANK WEIGHT was calculated in order to measure tournament level, since we divide sum of player's ranking by the draw size; again, the higher the number, the lower the level.

The second group of Data Set, Marginal Payoff and Prize Money Data Set consists of: TOURNAMENT (a number given for each number, the maximum is 129), YEAR (what year tournament was played), PRIZE MONEY BREAKDOWN (the distribution of the monetary prize awarded to players in a tournament based on round), POINTS (number of points awarded to players in a tournament based on round), ROUND (sequence of matches played, from 1 to 7), MARGINAL PAYOFF (discounted sum of the differences in spread) TOTAL PRIZE MONEY SINGLES (total prize money available to the singles tournament), and WINNERS PRIZE MONEY (total prize money given to the champion of the tournament).

Most variables in the second group of the Data Set were created for empirical purposes. MARGINAL PAYOFF is a variable representing the discounted sum of the differences between two prizes based on ROUND. When calculating players' winning probabilities for MARGINAL PAYOFF I decided to go with Rosen's approach and assign a probability of 1/2 to each player in all rounds. Hence, MARGINAL PAYOFF for the first round in a 32 draw tournament is derived as: (p2-p1)+(p3-p2)*0.5+(p4-p3)*0.5^2+(p5-p4)*0.5^3+(p6-p5)*0.5^4, where pi stands for prize money in round i. MARGINAL PAYOFF for the second round in the same draw size is: (p3-p2)+(p4-p3)*0.5+(p5-p4)*0.5^2+(p6-p5)*0.5^3. Remaining MARGINAL PAYOFFS can be calculated similarly. Because total purse normally includes doubles prize money, and since this study is only analyzing singles matches, I have decided to create the variable TOTAL PRIZE MONEY SINGLES, which can be calculated as: (16*p1)+(8*p2)+(4*p3)+(2*p4)+(1*p5)+(1*p6). This is the total prize money for a 32-draw size tournament, where pi stands for prize money in round i.

HYPOTHESES

Tournament theory, examined by Lazear and Rosen (1981) and later by Rosen (1986), has been used to describe certain situations where differences in wage are not based on the marginal productivity of an individual but instead based upon relative differences between those. Many economists used Rosen's theory to explain why people like Jeffrey Immelt and Warren Buffet are "overpaid". It might be because Immelt and Buffet are not paid for the work they do but, rather, to inspire workers in their firm. In other words, "normal" employees suffer in "underpaying" jobs hoping that one day they will become "overpaid" CEOs like Immelt and Buffet. That's what we call tournament theory. Promotion on the professional tennis tour is similar to the one in the office as great players are promoted, and less capable players are not. Roger Federer, for instance, is not paid to try his best, or to play perfect tennis, but he is paid to beat the rest of the players in the tour (which might be enough to get Federer's best out of him, most of the times).

When applying Rosen's theory to professional tennis, one is expected to see an increase in effort level as spread level between two prizes increases. However, one has to be sensitive with "how large" of an increase, since the bigger the spread, more effort will be required, which means higher marginal cost supplied by players. Hence, there must be an equilibrium level to be found in the spread difference. Not only that, but tournament organizers ought to know that more skilled players will choose to play tournament where spread level is wider. Rosen points out that the level of effort will be directly affected by differences in spread level. Ivankovic (1995) tested such hypothesis and found it to be true. As spread level increases players effort level will increase as well. However, it is important to make a difference between total prize money (total purse) and prize money breakdown (spread level). Total prize money might not affect effort level straightforwardly, if spread level is badly distributed for instance. Rosen also points out that 50% of total prize money is awarded to the last four players in a tournament. Ivankovic also tested such theory and found out that most tournaments offer around 40% of total prize money to the last four players.

I have structured my hypotheses based on Rosen's tournament theory, and Ivankovic's and Dinterman's empirical studies. Hence, the goal of this thesis is to test the following hypotheses:

Hypothesis 1

> Marginal Payoff (spread) is expected to increase at a linear or at an increasing rate up to the final round, and then have a distinct jump in the final match. Spread difference in the final round is expected to be the highest when compared to earlier rounds.

Hypothesis 2

> Best 4 players of a tournament (champion, finalist, and 2 semi-finalists) are expected to receive 50% of total prize money.

Hypothesis 3

> The probability a player wins a match increases if he wins the first set. I expect to see the effect of winning first set being more important to lower ranked players.

Hypothesis 4

Upsets (when the lower ranked player wins the match) are more likely to occur in ATP tournaments (played in best of 3 sets) when compared to Grand Slam tournaments (played in best of 5 sets). An increase in number of sets played will increase the probability for the better player to win a match.

Hypothesis 5

A change in spread level is expected to directly influence the outcome of a match. As spread level increases I expect to see more wins coming from lower skilled players.

Hypothesis 6

Total prize money will not affect outcome of a match but entry in a tournament. I expect to see better players entering tournaments where total purse is higher.

Hypothesis 7

I expect to see that the effect of outcome depends on players' ranking levels. Less competition is expected to be seen when ranking difference gets wider.

RESULTS

Descriptive Statistics

The dataset contains observations on 5,279 matches in 129 tournaments played in the 2007-2009 period. Table I provides the reader with tournament statistics on means, standard deviations, minimum, and maximum values of all variables used in the regressions.

Table I. Means, standard deviations, minimum, and maximum values for regression variables. (n= 5,279)

Variable	Mean	Std. Dev.	Minimum	Maximum
Tournament	77.19	40.36	1	129
Year	2008.45	0.53	2007	2009
Sets	0.19	0.39	0	1
Round	1.93	1.16	1	7
Draw	57.65	37.19	28	128
Match	19.16	17.16	1	64
Top Ranking	38.59	36.75	1	557
Bottom Ranking	116.66	167.52	2	2000
First Set Winner	0.83	0.46	0	1
Outcome	0.67	0.46	0	1
Surface	1.55	0.71	1	4
Rank Difference	78.05	157.04	0	1999
Rank Weight	7874.59	4903.08	1869	18797

Average Ranking Weight	149.41	60.02	45.25	361.42
Prize Money Breakdown	21956.25	41122.44	1000	800000
Points	38.01	75.86	0	1200
Marginal Payoff	61789.85	81530.04	8836.17	850000
Total Prize Money Singles	2058177	2429653	272280	7547580
Winners Prize Money	443060.7	475142.1	64000	1600000
Maximum Rounds	5.642167	0.78	5	7
Final	0.024	0.15	0	1
Semi Final	0.04	0.21	0	1
Quarter Final	0.09	0.29	0	1
Top 16	0.19	0.39	0	1
Top 32	0.38	0.48	0	1
Top 64	0.15	0.36	0	1
Top 128	0.09	0.29	0	1
Hard Court	0.57	0.49	0	1
Clay Court	0.30	0.46	0	1
Grass Court	0.11	0.31	0	1
Carpet Court	0.005	0.07	0	1

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

The spread level or marginal payoff was obtained through Rosen's derivation. According to Rosen, the incentive to win a round is determined by the discounted sum of differences in prize money breakdown. As Ivankovic mentioned in his thesis, marginal payoff = (expected price due to advancement – loser's pay). One important thing about marginal payoff is the winning probability assigned to each player in each round. I have set same probabilities to all players (1/2) for all rounds but we know that 150th ranked player has less than 1/2 probability of beating Rafael Nadal for instance, who is currently 3rd in the world. Assuming 1/2 probability for all matches might produce biased coefficients for marginal payoff in the first rounds, but as the number of rounds to be played in a tournament decrease, probability is expected to converge to ½. Further research in how to set winning probabilities might be useful for in this case.

Variables that influence the outcome of a match, from Table I, ought to be mentioned. SETS ranges from 0 to 1, being 0 if played best of 3 sets and 1 if played best of 5 sets. Its mean is 0.19, which tells us that players played more matches determined in best of 3 sets during the 2008 and 2009 seasons. Variable TOP RANKING ranges from 1 to 557, with a mean of 38. This is interesting because even though the maximum value for TOP RANKING is considered to be a very low professional ranking (557), one can see that TOP RANKING values were quite close to its mean (standard deviation is about 36), indicating that not many TOP RANKING players had really low rankings, which can be explained by the fact that two qualifiers might play each other in the first round. When looking at BOTTOM RANKING one can observe a mean of 116, minimum of 2, and maximum of 2000. Most of players that received a WC, an invitation to participate in a tournament, were not ranked within the ATP norms. For those players, I decided to give a ranking equal 2000, so one I am able to differ them from the rest. When doing a descriptive statistics of BOTTOM RANKING excluding all rankings equal 2000, I see that the maximum value for BOTTOM RANKING will equal 1734, which is not much different than 2000. We also observe that about 15 players did not possess an ATP ranking in this data set. Standard deviation for BOTTOM RANKING is equal to 167, which means that BOTTOM RANKING is more spread out over a large range of values when compared to TOP RANKING.

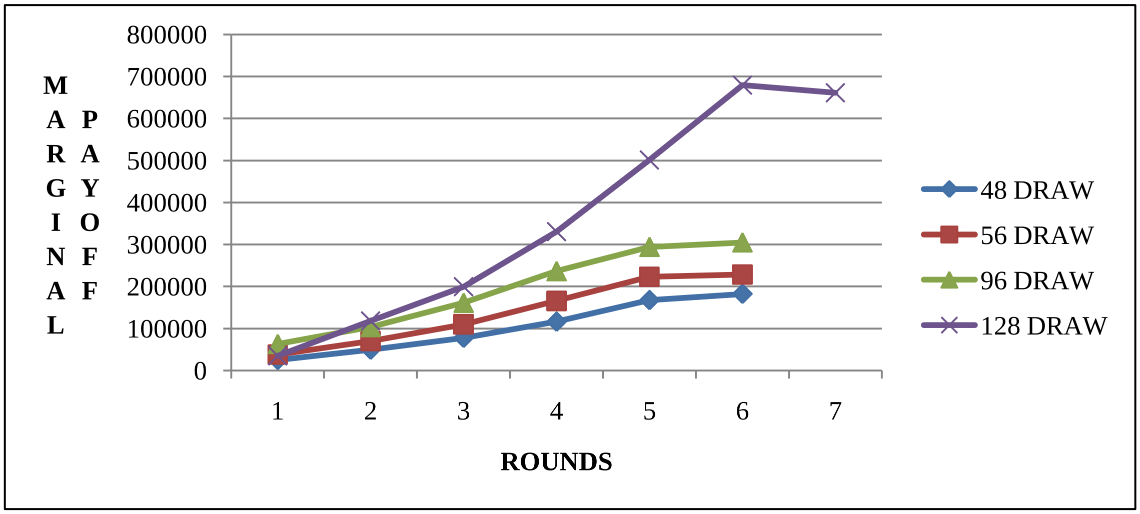
Variable SURFACE was divided into four different categories: 1 if surface is hard court, 2 if clay court, 3 if grass court, and 4 if carpet. From Table 1, one can see that most tournaments during 2008 and 2009 were played on hard court surface (about 57%). The next most played surface was clay court (30%) followed by grass court (11%). The remaining 2% of matches were played on carpet court surface. TOTAL PRIZE MONEY SINGLES has a mean of $2,058.177. Minimum purse distributed in a tournament equaled $272,280 and maximum purse distributed was $7,547,580. On the other side, MARGINAL PAYOFF has a mean equal to $61,789.85. Lowest level of spread offered in a tournament was $8,836.17 and highest level of spread offered was $850,000.

Test of Hypothesis 1

Figures III through VIII provide graphs for MARGINAL PAYOFF levels and PRIZE MONEY BREAKDOWN levels, where ROUNDS is the number of rounds played in a tournament. There are five rounds in a tournament that consists of 28, and 32 players, six rounds in a 48, 56, and 96 players tournament, and seven rounds for a 128 players tournament.

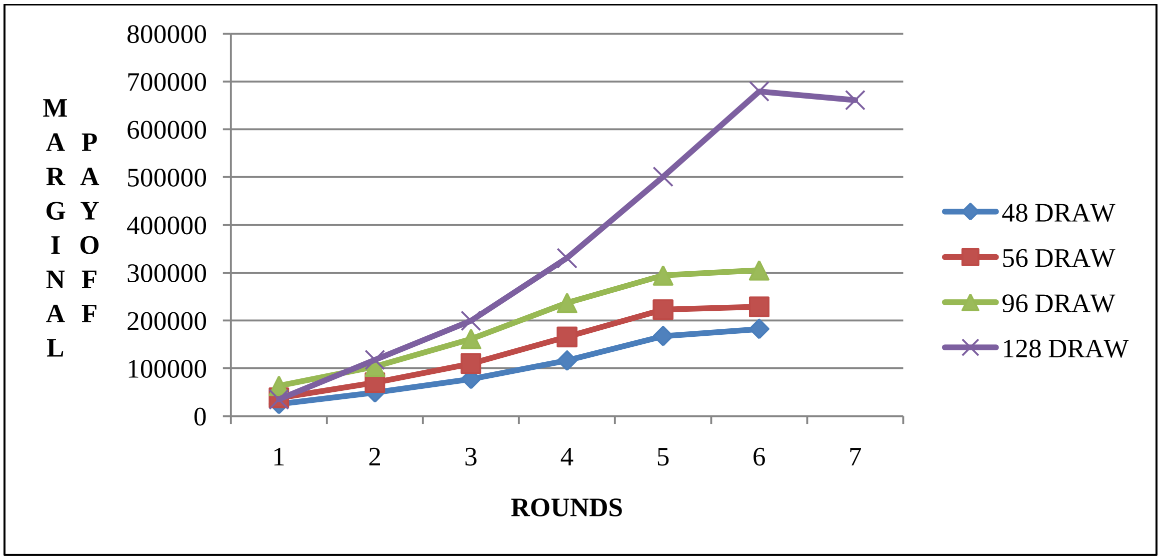
As we move towards the final round the levels of MARGINAL PAYOFF and PRIZE MONEY BREAKDOWN are expected to increase, at a linear or increasing rate, as the amount of rounds left to be played in a tournament decreases; however, in the final round, levels of MARGINAL PAYOFF are expected to have a distinct jump. According to Ivankovic (1995), theory suggests that the amount of increase in spread will depend on risk preference adopted by players. A risk neutral player will be satisfied with a constant spread up to the semi-final match, while in the final a larger increase in spread is required. If the player is risk averse, then spread needs to be increasing at an increasing rate with an even larger difference in spread between the champion and finalist. Spread needs to be increasing in order to maintain risk averse players interested in entering the tournament.

Figure III: Mean Leves for MARGINAL PAYOFF when i=ROUND and DRAW= 28 and 32



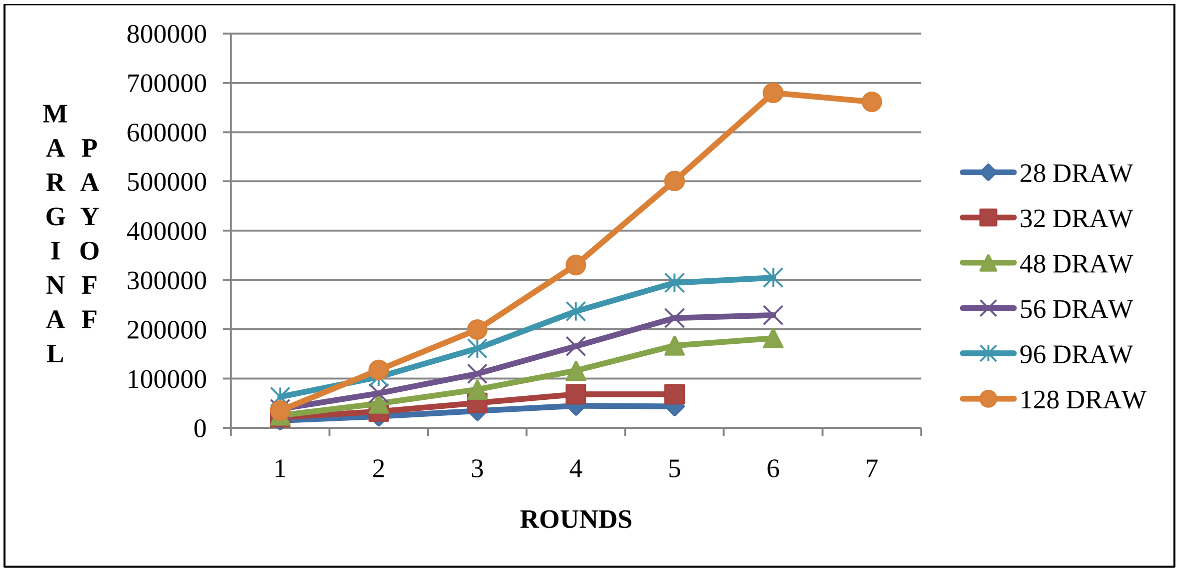
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure IV: Mean levels for MARGINAL PAYOFF when i=ROUND and DRAW=49, 56, 96, AND 128



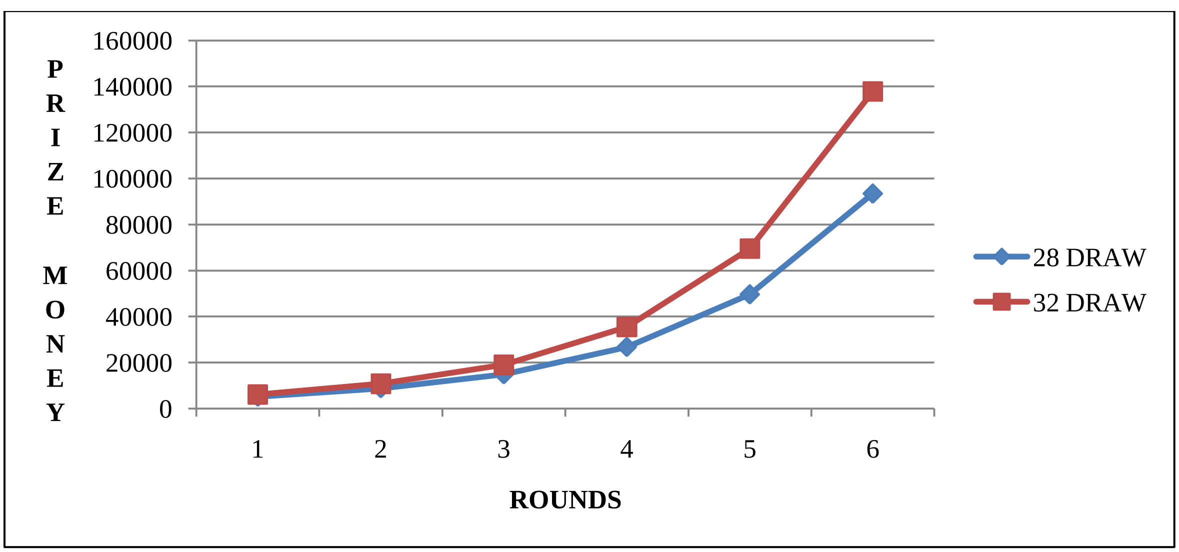
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure V: Mean Leves for MARGINAL PAYOFF when i=ROUND and DRAW=28, 32, 48, 56, 96, and 128



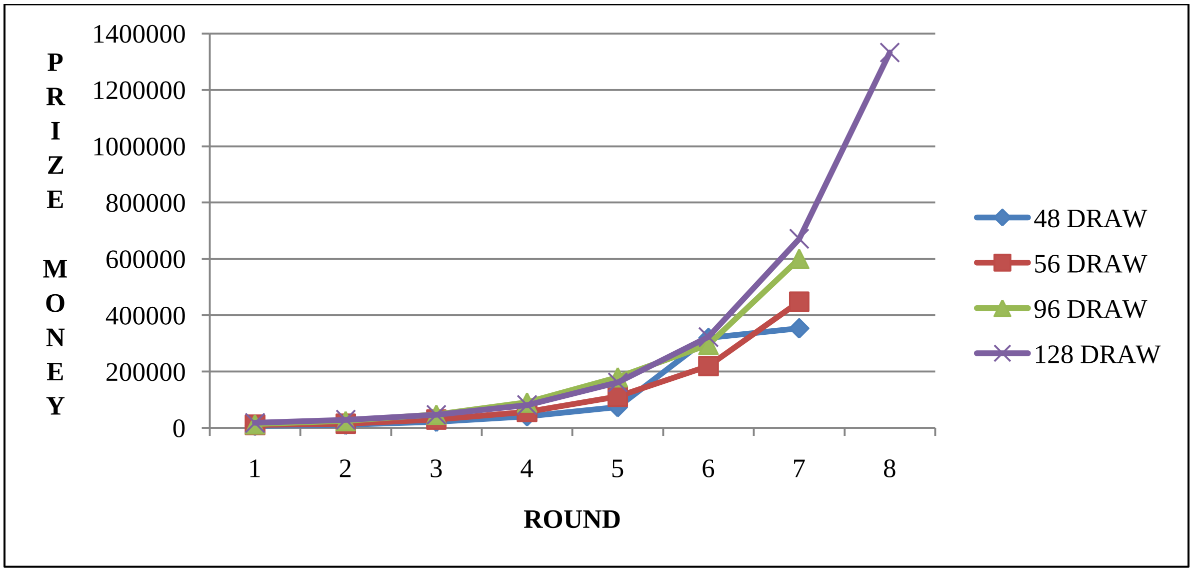
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure VI: Mean Levels for PRIZE MONEY BREAKDOWN when i=PRIZE and DRAW=28 and 32



Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure VII: Mean Levels for PRIZE MONEY BREAKDOWN when i= PRIZE and DRAW= 48, 56, 96, and 128

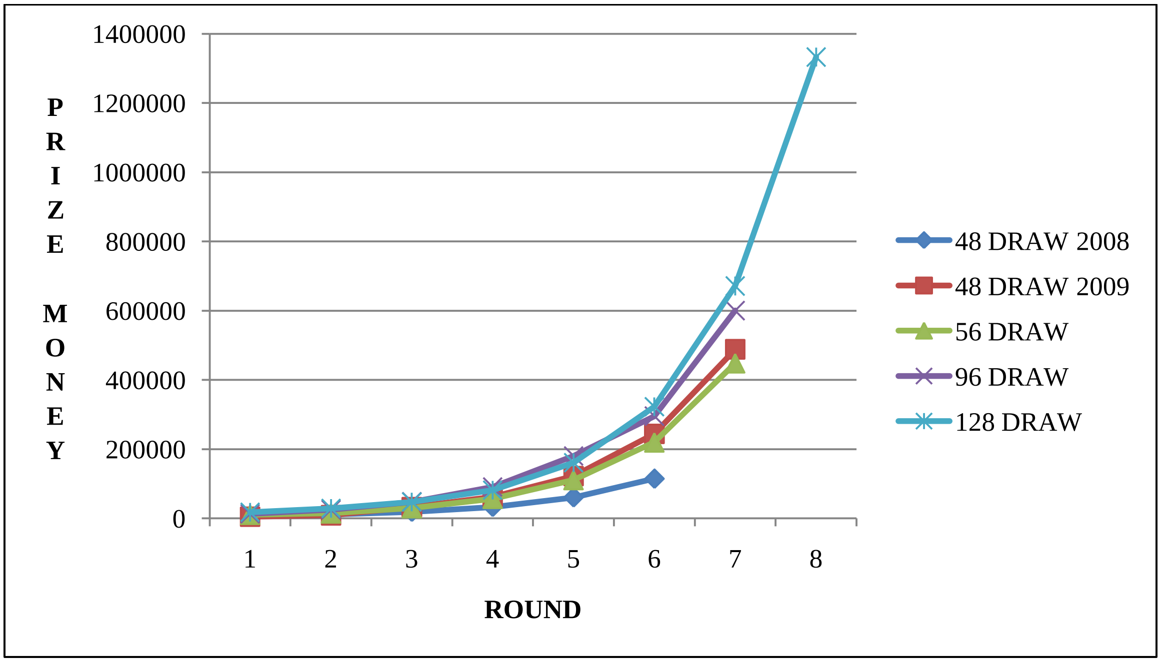


Source: Author's calculations from data gathered on the ATP Tour website, 2010.

When studying figures III to V we conclude that tournament theory does not necessarily apply in the reality of professional tennis tournaments. Marginal payoffs between rounds are increasing, but at a decreasing rate (percentage change between spread levels is decreasing). In all draws we see a distinct fall in marginal payoff in the final round, but in the 28 and 128 draw format we actually observe a negative number, meaning that the marginal payoff is lower in the final round compared to the semi-final.

Observations from figures VI to VIII show different results. Such tables provide the reader with changes between prize money instead of marginal payoff. Most draw formats show similar results in prize money difference with an exception of the 48 draw format, which shows a large increase of prize money in round 5 and an even larger decrease in the last round. This result could be explained by the fact that in 2008 two out of the four 48 draw tournaments distributed only six prizes money, meaning that two tournaments were played in a total of five rounds instead of six. To test if such assumption is true I decided to separate prize money differences by year, round, and match. The results can be seen on figure VIII.

Figure VIII: Mean Levels for PRIZE MONEY BREAKDOWN when i=ROUND, DRAW=48 and YEAR=2008, DRAW=48 and YEAR=2009, and DRAW=56, 96, and 128



Source: Author's calculations from data gathered on the ATP Tour website, 2010.

When splitting observations for the 48 draw tournaments by year, round, and match we see that prize money increases at a constant rate (it doubles every round). We can also see a large difference in percentage between Prize 1 and Prize 2, only in tournaments played in 2009.[4] Overall, we are able to conclude that PRIZE MONEY BREAKDOWN increases at a relatively constant rate, and it seems to have a small increase in the last round.

From above all graphs, MARGINAL PAYOFF results for 28 and 128 draw tournaments increases at a decreasing rate up to the final match, and then it drops. MARGINAL PAYOFF will also increase at a decreasing rate for all other types of tournaments, but it will not become negative in the last round, as it will increase at a much lower rate, almost equal to zero at times.

Changes in PRIZE MONEY BREAKDOWN are rather linear across rounds for all types of tournaments. Observing the level of PRIZE MONEY BREAKDOWN in the last round it is obvious to see that there is no distinct jump, it actually has a very small increase.

From the above graphs and tables located in the appendix, most results lead towards the rejection of the first hypothesis. Ivankovic (1995) measured same hypothesis and found similar results. Comparing my results with his, the only notable difference is that Ivankovic found that MARGINAL PAYOFF drops and it actually becomes negative, in the last round, for all types of tournaments. I can conclude that the structure of the distribution of prize money in professional tennis tournaments does not follow the theory; however it seems to be changing in a way that favors tournament model.

[4] See Appendix for more details.

Test of Hypothesis 2

Hypothesis 2 will address the assumption that top four players in a tournament (champion, finalist, and two semi-finalist) receive 50% of total prize money. Second hypothesis supports the idea that spread level is expected to increase at a linear or increasing rate as the number of rounds to be played in a tournament decrease (hypothesis 1).

I observed three different levels of prize distribution (champion, finalist, and semi-finalist) and compared them with total prize money distributed in the tournament. Total prize money is only measured for the singles tournament because that's the only data I am examining. Table IX summarizes the results found.

Table XVII. Means, TOTAL PRIZE MONEY SINGLES, and % OF TOTAL PRIZE MONEY SINGLES, where DRAW = 28, 32, 48, 56, 96, and 128.

Variable	Mean	Variable	Mean
Winner Prize 28	93417.08	Winner Prize 32	137868.2
Finalist Prize 28	49594.31	Finalist Prize 32	69513.87
Semi Prize 28	26731.22	Semi Prize 32	35490.24
Total Purse 28	387473.3	Total Purse 32	538041
% Total Purse	**43.8**	**% Total Purse**	**45.140112**
Variable	Mean	Variable	Mean
Winner Prize 48	353693.3	Winner Prize 56	448335.7
Finalist Prize 48	319522.9	Finalist Prize 56	219708.8
Semi Prize 48	74030.71	Semi Prize 56	111036.7

Total Purse 48	1485727	Total Purse 56	1850889
% Total Purse	**50.29**	**% Total Purse**	**42.09**
Variable	Mean	Variable	Mean
Winner Prize 96	600333.3	Winner Prize 128	1332735
Finalist Prize 96	295333.3	Finalist Prize 128	671430
Semi Prize 96	180300	Semi Prize 128	322590
Total Purse 96	2321013	Total Purse 128	6773250
% Total Purse	**46.35**	**% Total Purse**	**34.35**
	Avg. % Total Purse	**43.67**	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

From table XVII we can easily see only the 48 draw format distributes 50% of total purse to the top four players. One can observe that most tournaments organized by the ATP do distribute almost half of total purse to the top four players (with an exception of Grand Slams), with an average distribution of total purse equal to 43.67%. It is interesting to see that in the 128 draw tournaments the top four players receive much less in terms of percentage (total purse) when compared to the rest of draws. An explanation could be that Grand Slam tournaments have a lot more players in the draw and that there is no need to create a larger incentive for the last four players to win their matches since they are in the most important tournaments played on the ATP, and every player on the tour dreams about winning a title.

Ivankovic tested same hypothesis on his work and found that most tournaments offer around 40% of the total purse to the top four players. He also pointed out that Grand

Slam tournaments offered an even lower percentage amount, which in his study equaled 31.95%. Thus, I can conclude that, again, the professional tennis tour is not applying Rosen's theory when organizing tournaments. Thus, results lead me to reject hypothesis 2. However, I can also conclude that in the past two seasons (2008 and 2009) tournaments have been structured in a different way when compared to 1995, and now it seems that the ATP Tour is moving its prize distribution towards Rosen's theory.

Test of Hypothesis 3

In hypothesis 3 I will test if the probability of winning a match increases after a player wins the first set. I expect to see the effect of winning the first set being more important to lower ranked players. Tables XVIII-XV will provide results.

Table XVIII. Percentages for FIRST SET WINNERi, and OUTCOMEi, where i=0 or 1.

First Set Winner	Outcome	Percentage
1	0	30.62
1	1	69.38
0	0	39.32
0	1	60.68

Note: FIRST SET WINNER=1 means that better player won the first set, and 0 that better player lost the first set. OUTCOME=1 means that better player won the match, and 0 that better player lost the match.
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XIX. Percentages for FIRST SET WINNERi where i=0 or 1.

First Set Winner	Percentage
0	19.09
1	79.64

Note: FIRST SET WINNER=1 means that better player won the first set, and 0 that better player lost the first set.
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XX. Percentages for OUTCOMEi where i=0 or 1.

Outcome	Percentage
0	32.59
1	67.41

Note: OUTCOME=1 means that better player won the match, and 0 that better player lost the match.
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

From Table XVIII I observe the following:

- 30.62% of times the better player won the first set but lost the match.
- 69.38% of times the better player won the first set and also won the match.
- 39.32% of times the better player lost the first set and also lost the match.
- 60.68% of times the better player lost the first set but still won the match.

With that, I can conclude that if the worse player wins the first set, his chances of winning the match increase by 8.7%, which will be the same percentage for the better player. From table XIX I see that the worse player won the first set 19.09% of times, and that the better player won the first set 79.64% of times. From table XX I conclude that only 32.59% of matches were won by worse players versus 67.41% of matches won by better players. Finally, I am able to accept the hypothesis that winning probability goes up after a player wins the first set. However, I am not able to conclude that the first set is more important for worse players. Although, I can argue that going from 30.6% to 39.3% is a 1/3 improvement, while going from 60.7% to 69.4% is only a 1/6 improvement.

Regression Results

Tests of Hypotheses 4 and 5

Assume A and B are individuals trying to hit one serve at a target on the tennis court. With that being said, we all ought to agree both of them have only one chance to hit the target. Now, assume individual A has comparative advantage in serving when compared to individual B. Increasing the number of serves will increase the chances of both individuals hitting the target, meaning that the probability of hitting a target with one chance is lower than if there were two chances, for instance. However, the point here is that as I increase the number of chances given to both individuals I expect to see individual A hit the target more times than individual B. Thus, the assumption behind this idea is that winning probability of individual A, regarding the serve contests, is expected to increase every time I give both players an extra chance to hit the target.

Hypothesis 4 will test if upsets (when the lower ranked player wins the match) are more likely to occur in ATP tournaments (played in best of 3 sets) when compared to Grand Slam tournaments (played in best of 5 sets). An increase in number of sets played is expected to increase the probability for the better player to win a match.

Hypothesis 5 will analyze the effects of spread on outcome. An increase in spread level is expected to directly influence the outcome of a match (I expect to see more wins coming from lower skilled players because I believe tournament prize money is their primary source of income).

I estimated the effects of SETS and MARGINAL PAYOFF on the dependent variable OUTCOME. Hence, I constructed the following regression:

$$\begin{aligned} Outcome_{tb} &= \alpha_0 + \alpha_1 TopRanking_i + \alpha_2 BottomRanking_i + \alpha_3 ClayCourt_i \\ &+ \alpha_4 GrassCourt_i + \alpha_5 CarpetCourt_i + \alpha_6 MarginalPayoff_i 10k \\ &+ \alpha_7 Sets_i + \alpha_8 TopRankingGS_i + \alpha_9 BottomRankingGS_i + \mu_{tb} \end{aligned} \tag{17}$$

$Outcome_{tb}$ is the result of a match i played between t (top) and b (bottom) players. $TopRanking_i$ is the ranking of better player when match i was played. $BottomRanking_i$ is the ranking of worse player when match i was played. $ClayCourt_i$ stands for the type of surface match i was played, which is 1 if clay and 0 otherwise. $GrassCourt_i$ also stands for the type of surface match i was played, which is 1 if grass and 0 oherwise. $CarpetCourt_i$ is surface that match i was played, which is 1 for carpet and 0 otherwise. The omitted surface variable is hard court. Variable $MarginalPayoff_i 10k$ is the marginal prize given for advancing to the next round in a match i, $Sets_i$ is the format match i was played, either best of 3 sets of best 5 sets. $Sets_i$ takes on dummy variables of 1 if played best of 5 sets, and 0 if played best of 3 sets. $TopRankingGS_i$ is the ranking of better when match i was played in Grand Slams. $BottomRankingGS_i$ is the ranking of worse player when match i was played in Grand Slams. Let's focus on results in table XXI.

Table XXI. Marginal effects after logit regression on SETS and MARGINAL PAYOFF for the 2007-09 ATP Tour. Dependent variable= OUTCOME. Absolute z values in parenthesis.

Variable	(1)
Top Ranking	-.0023286 (9.97)
Bottom Ranking	.0004775 (6.56)
Clay Court	.0115025 (0.80)
Grass Court	.0153684 (0.71)
Carpet Court	-.0898073 (1.01)
Marginal Payoff 10k	**-.0020323 (2.06)**
Sets	**.0827337 (2.67)**
Top Ranking Grand Slam	-.0024583 (3.89)
Bottom Ranking Grand Slam	.0008218 (2.54)
(*) dy/dx is for discrete change of	dummy variable

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XXI represents marginal effects after a logit regression from equation (17). An increase in top players' ranking (actually a decrease in variable Top Ranking because lower number means better ranking) will increase the amount of matches defeated. On the other side, a decrease in bottom players' ranking (actually an increase in variable Bottom Ranking) will decrease the amount of upsets. For example, the value of .011 shows that a defeat is 1% more likely to happen if a match is played on clay court, than to a match played on hard court. It also shows that a defeat is 1.5% more likely to happen if a match is played on grass court compared to a match played on hard court. On the other side, it yields that an upset is 8.9% more likely to happen if a match is played on carpet court, when compared to a match played on hard court.

Matches played in best of 3 sets do have a higher frequency of upsets, about 8% more, compared to matches played best of 5 sets. Hence, I am able to accept hypothesis 4, which assumed that an increase in number of sets played, would increase the probability for the better player to win a match.

As marginal payoff level increases, the number of upsets increases as well. An increase of $10,000 in marginal payoff will increase the frequency of upsets by .2%, which is very small. Perhaps a more significant result would be an increase of $100,000 in marginal payoff, which would increase upsets by 2%. Still, I am able to accept hypothesis 5, where an increase in spread level is associated with an increase in the number of upsets. A speculation to explain such result is that better players' primary source of income must come from sponsors, or somewhere else, but not come from tournaments' total purse. On the other side, perhaps an interesting topic for investigation is whether or not worse players' primary source of income comes from tournaments' total purse.

Last observation from table XXI yields that an increase in grand slams players' ranking will increase their number of wins. On the other side, a decrease in grand slams players' ranking will decrease their number of wins.

Test of Hypothesis 6

Hypothesis 6 will test the effect of total prize money on the entry level of a tournament. Theory suggests that total purse will have a positive relationship with ranking (better players are more likely to enter tournaments with higher purse levels).Hence, I expect to see that an increase in purse will increase quality of play in a tournament.

Table XXII. Effect of total purse on entry into tournaments for the 2007-09 ATP Tour. Dependent variable= AVERAGE RANK WEIGHT. Absolute t-statistics in parenthesis.

Variable	(1)
Total Prize Money 10K	**-.142503 (4.37)**
Adj. R^2	0.1306
F value	19.07

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Results on table XXII show that t statistics on total purse (4.37) is significant. Also, the coefficient for total purse is negative, which implies an opposite relationship with average rank weight. In other words, as the level of total purse increases, the level of players, in terms of ranking, entering a tournament increases. It is important to clarify, for uninformed readers, that higher average rank weight means lower quality of play. An increase of $10,000 in total prize money will decrease average rank weight by around 14 spots in the ATP ranking list, which is very significant. Finally, I conclude that tournament theory does predict, correctly, entry into an ATP Tour tournament.

Test of Hypothesis 7

Hypothesis 7 will address the relationship between outcome and players' ranking levels. Less competition is expected to be seen when ranking difference gets wider. Replacing ranking difference for top ranking and bottom ranking into equation (17) yields:

$$\begin{aligned}Outcome_{tb} &= \alpha_0 + \alpha_1 ClayCourt_i + \alpha_2 GrassCourt_i + \alpha_3 CarpetCourt_i \\ &+ \alpha_4 MarginalPayoff_i 10k + \alpha_5 Sets_i + \alpha_6 TopRankingGS_i \\ &+ \alpha_7 BottomRankingGS_i + \alpha_8 RankingDifference_i +_i \mu_{tb}\end{aligned} \tag{18}$$

where $RankingDifference_i$ is the absolute difference between $TopRanking_i$ and $BottomRanking_i$ in match i. Regression on tournament level is represented by table XXIII.

Table XXIII. Marginal effects after logit regression on RANKING DIFFERENCE for the 2007-09 ATP Tour. Dependent variable= OUTCOME. Absolute z values in parenthesis.

Variable	(1)
Clay Court	.0047745 (0.33)
Grass Court	.0027262 (0.13)
Carpet Court	-.0654743 (0.75)
Marginal Payoff 10k[5]	.00154 (1.67)
Sets	.0638548 (3.64)
Ranking Difference	**.0004622 (6.32)**
(*) dy/dx is for discrete change of dummy variable from 0 to 1	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

[5] This is the only case where Marginal Payoff 10k shows a positive coefficient sign.

From table XXIII I conclude that a spot increase in ranking difference decreases the number of upsets by .04% only. For example, a 10 spot increase in ranking difference decreases the number of upsets by a small amount of .4%. I can also conclude that for this model, variables $TopRanking_i$ and $BottomRanking_i$ are more efficient measures than $RankingDifference_i$ variable. Coefficient $M\arg inalPayoff_i$ became positive but not insignificant after ranking difference was added to the new regression. An increase of $10,000 in marginal payoff would increase defeats by .1%, which is extremely small. Variables $ClayCourt_i$ and $GrassCourt_i$ became insignificant. The effect of $Sets_i$ on outcome became stronger. Matches played in best of 3 sets continue to have a higher frequency of upsets, but now instead of about 8% more, is just 6% more when compared to matches played best of 5 sets. For curiosity, I decided to see what happens to outcome when we divide rankings, and better from weaker players in subgroups. So, I generated new variables in STATA that separated players and rankings as:

- Better players ranked in the top 10 versus worse players also ranked in the top 10
- Better players ranked in the top 10 versus worse players ranked in the top 11 to 24
- Better players ranked in the top 10 versus worse players ranked in the top 25 to 49
- Better players ranked in the top 10 versus worse players ranked in the top 50 to 74

- Better players ranked in the top 10 versus worse players ranked in the top 75 to 100
- Better players ranked in the top 10 versus worse players ranked worse than 100

- Better players ranked in the top 11 to 24 versus worse players ranked in the top 11 to 24
- Better players ranked in the top 11 to 24 versus worse players ranked in the top 25 to 49
- Better players ranked in the top 11 to 24 versus worse players ranked in the top 50 to 74
- Better players ranked in the top 11 to 24 versus worse players ranked in the top 75 to 100
- Better players ranked in the top 11 to 24 versus worse players ranked worse than 100

- Better players ranked in the top 24 to 49 versus worse players ranked in the top 24 to 49
- Better players ranked in the top 24 to 49 versus worse players ranked in the top 50 to 74
- Better players ranked in the top 24 to 49 versus worse players ranked in the top 75 to 100

- Better players ranked in the top 24 to 49 versus worse players ranked worse than 100

- Better players ranked in the top 50 to 74 versus worse players ranked in the top 50 to 74
- Better players ranked in the top 50 to 74 versus worse players ranked in the top 75 to 100
- Better players ranked in the top 50 to 74 versus worse players ranked worse than 100

- Better players ranked in the top 75 to 100 versus worse players ranked in the top 75 to 100

- Better players ranked in the top 75 to 100 versus worse players ranked worse than 100.

Adding all new variables into equation (18) gives:

$$\begin{aligned}
Outcome_{tb} &= \alpha_0 + \alpha_1 ClayCourt_i + \alpha_2 GrassCourt_i + \alpha_3 CarpetCourt_i \\
&+ \alpha_4 M\arg inalPayoff_i 10k + \alpha_5 Sets_i + \alpha_6 Top10vsBot10 + \alpha_7 Top10vsBot25 \\
&+ \alpha_8 Top10vsBot50 + \alpha_9 Top10vsBot75 + \alpha_{10} Top10vsBot100 \\
&+ \alpha_{11} Top10vsBot > 100 + \alpha_{12} Top25vsBot25 + \alpha_{13} Top25vsBot50 \\
&+ \alpha_{14} Top25vsBot75 + \alpha_{15} Top25vsBot100 + \alpha_{16} Top25vsBot > 100 \\
&+ \alpha_{17} Top50vsBot50 + \alpha_{18} Top50vsBot50 + \alpha_{19} Top50vsBot75 \\
&+ \alpha_{20} Top50vsBot100 + \alpha_{21} Top50vsBot > 100 + \alpha_{22} Top75vsBot75 \\
&+ \alpha_{23} Top75vsBot100 + \alpha_{24} Top75vsBot > 100 \\
&+ \alpha_{25} Top100vsBot100 + \alpha_{26} Top100vsBot > 100 \\
&+ \alpha_{27} Top > 100vsBot > 100 + \mu_{tb}
\end{aligned} \tag{19}$$

where, $Top10vsBot10$ is when better players ranked in the top 10 play against worse players also ranked in the top 10. $Top10vsBot25$ is when better players ranked in the top 10 play against worse players ranked between 11 to 25. $Top10vsBot50$ is when better players ranked in the top 10 playing against worse players ranked between 26 to 50. $Top10vsBot75$ variable stands for better players ranked in the top 10 versus worse players ranked between 51 to 75. $Top10vsBot100$ variable stands for better players ranked in the top 10 versus worse players ranked between 76 to 100. $Top10vsBot > 100$ stands for better players ranked in the top 10 versus worse players ranked poorer than 100. $Top25vsBot25$ variable stands for better players ranked between 11 to 25 versus worse players ranked between 11 to 25 as well. $Top25vsBot50$ is when better players ranked between 11 to 25 play against weaker players ranked between 26 to 50. $Top25vsBot75$ variable describes the situation when better players ranked between 11 to 25 play versus weaker players ranked between 51 to 75. $Top25vsBot100$ variable describes the situation when better players ranked between 11 to 25 play versus weaker players ranked between 76 to 101. $Top25vsBot>100$ stands for better players ranked between 25 to 50 versus worse players ranked poorer than 100. $Top50vsBot50$ variable describes the situation when better players ranked between 26 to 50 play versus weaker players also ranked between 26 to 50. $Top50vsBot75$ variable is explained by matches in which better players ranked between 26 to 50 play versus weaker players ranked between 51 to 75. $Top50vsBot100$ variable is explained by matches in which better players ranked between 26 to 50 play versus weaker players ranked between 76 to 101. $Top50vsBot>100$ variable is explained by matches in which better players ranked between 26 to 50 play versus weaker players ranked poorer than 10o. $Top75vsBot75$ is when better players ranked between 51 to 75 play against worse players also ranked between 51 to 75. $Top75vsBot100$ is when better players ranked between 51 to 75 play against worse players ranked between 76 to 101. $Top75vsBot>100$ is when better players ranked between 51 to 75 play against worse players ranked poorer than 100. $Top100vsBot100$ is when better players ranked between 76 to 101 play against worse players also ranked between 76 to 101. $Top100vsBot>100$ is when better players ranked between 76 to 101 play against worse players ranked poorer than 100. Lastly, $Top>100vsBot>100$ is when better players ranked poorer than 100 play against worse players ranked poorer than 100. Regression on players' ranking subgroup is represented by table XXIV.

Table XXIV. Marginal effects after logit regression on PLAYERS' RANKING SUBGROUP for the 2007-09 ATP Tour. Dependent variable= OUTCOME. Absolute z values in parenthesis.

Variable	(1)
Clay Court	.0120572 (0.83)
Grass Court	.0142309 (0.66)
Carpet Court	-.0683389 (0.77)
Marginal Payoff 10k	-.0026491 (2.26)
Sets	.0563614 (3.03)
Top 10 versus Bottom 11 to 25	**.0815007 (1.75)**
Top 10 versus Bottom 26 to 50	**.1089054 (2.45)**
Top 10 versus Bottom 51 to 75	**.1580663 (3.83)**
Top 10 versus Bottom 76 to 101	**.2113679 (5.65)**
Top 10 versus Bottom > 100	**.2092679 (5.80)**
Top 11 to 25 versus Bottom 11 to 25	**-.0946906 (1.30)**
Top 11 to 25 versus Bottom 26 to 50	**-.0484663 (0.83)**
Top 11 to 25 versus Bottom 51 to 75	**.009893 (0.17)**
Top 11 to 25 versus Bottom 75 to 101	**.0948124 (1.83)**
Top 11 to 25 versus Bottom > 100	**.1207259 (2.60)**
Top 26 to 50 versus Bottom 26 to 50	**-.0782457 (1.22)**
Top 26 to 50 versus Bottom 51 to 75	**-.1055609 (1.71)**
Top 26 to 50 versus Bottom 76 to 101	**-.0160428 (0.27)**

Top 26 to 50 versus Bottom > 100	**.0019647 (0.03)**
Top 51 to 75 versus Bottom 51 to 75	**-.1664179 (2.36)**
Top 51 to 75 versus Bottom 76 to 101	**-.1636106 (2.42)**
Top 51 to 75 versus Bottom > 100	**-.0745935 (1.19)**
Top 76 to 101 versus Bottom 76 to 101	**-.1351779 (1.81)**
Top 76 to 101 versus Bottom > 100	**-.106418 (1.65)**
Top > 100 versus Bottom > 100	**-.0977171 (1.49)**
(*) dy/dx is for discrete change of dummy variable 1	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Results in table XXIV show that coefficient for $MarginalPayoff_i$ became negative again, and a $10,000 increase in marginal payoff will increase upsets by 2%. Variable $GrassCourt_i$ remained insignificant. The effect of $Sets_i$ on outcome became weaker when compared to table XXIII. Matches played in best of 3 sets continue to have a higher frequency of upsets, but now instead of about 6% more, is just 5% more when compared to matches played best of 5 sets. Variables $Top25vsBot75$, $Top50vsBot100$ and $Top50vsBot>100$ are statistically insignificant, meaning that matches played in these ranking intervals are considered to be competitive. Variable $GrassCourt_i$ also became insignificant. Coefficient $Top10vsBot10$is the omitted variable in the regression. Comparing $Top10vsBot10$ with $Top10vsBot25$I conclude that there are 8% more chances for the better player to win if he plays an opponent ranked between 11 to 25 instead of a top 10 player. When looking at $Top10vsBot50$ the probability that the top 10 ranked better player wins a match against worse player ranked between 26 to 50 increases even more, about 11%, when compared to a (better) top 10 player against another (worse) top 10 player. The probability that the top 10 ranked better player wins a match against a worse player ranked between 51 to 75 increases by about 16%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the top 10 ranked better player wins a match against a worse player ranked between 76 to 101 increases by about 21%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the top 10 ranked better player wins a match against a worse player ranked poorer than 100 also increases by about 21%, compared to a match where both (better and worse ranked) players are ranked in the top 10.

The probability that the better player, ranked between 11 to 25 wins a match against a worse player, also ranked between 11 to 25 decreases by about 9.5%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the better player, ranked between 11 to 25 wins a match against a worse player, ranked between 26 to 50 decreases by about 5%, compared to a match where both (better and worse ranked) players are ranked in the top 10. I cannot provide the probability that the better player, ranked between 11 to 25 wins a match against a worse player, ranked between 51 to 75 because coefficient is statistically insignificant. Hence, I can assume such matches are very competitive. The probability that the better player, ranked between 11 to 25 wins a match against a worse player, ranked between 76 to 101 increases by about 10%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the better player, ranked between 11 to 25 wins a match against a worse player, ranked poorer than 100 increases by about 12%, compared to a match where both (better and worse ranked) players are ranked in the top 10.

The probability that the better player, ranked between 26 to 50 wins a match against a worse player, also ranked between 26 to 50 decreases by about 8%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the better player, ranked between 26 to 50 wins a match against a worse player, ranked between 51 to 75 decreases by about 10.5%, compared to a match where both (better and worse ranked) players are ranked in the top 10. I cannot provide the probability that the better player, ranked between 26 to 50 wins a match against a worse player, ranked between 76 to 101 because coefficient is statistically insignificant. I also cannot provide the probability that the better player, ranked between 26 to 50 wins a match against a worse player, ranked poorer than 100 because again, coefficient is statistically insignificant. Hence, I can assume all these matches are very competitive.

The probability that the better player, ranked between 51 to 75 wins a match against a worse player, also ranked between 51 to 75 decreases by about 16%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the better player, ranked between 51 to 75 wins a match against a worse player, ranked between 76 to 101 also decreases by about 16%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the better player, ranked between 51 to 75 wins a match against a worse player, ranked poorer than 100 decreases by about 7%, compared to a match where both (better and worse ranked) players are ranked in the top 10.

The probability that the better player, ranked between 76 to 101 wins a match against a worse player, also ranked between 76 to 101 decreases by about 13%, compared to a match where both (better and worse ranked) players are ranked in the top 10. The probability that the better player, ranked between 76 to 101 wins a match against a worse player, ranked poorer than 100 decreases by about 10%, compared to a match where both (better and worse ranked) players are ranked in the top 10. Finally, The probability that the better player, ranked poorer than 100 wins a match against a worse player, also ranked poorer than 100 decreases by about 10%, compared to a match where both (better and worse ranked) players are ranked in the top 10. After observing most of results, I can accept the hypothesis that less competition is seen as differences in ranking increase. I can also conclude that there is a dominance for the better player ranked in the top 10 when compared to the rest of better ranked players. If we observe matches played between two top 10 ranked players, and compared with all other matches in which both (better and worse ranked) player were in the same ranking subgroup (11 to 25; 26 to 50; 51 to 75; 76 to 101; and >100) we can see a negative coefficient for such variables, meaning that chances for the better player to win a match goes down.

CONCLUSION

The tournament theory is a hot topic regarding pay structure in professional sports, and at the firm level. Many attempted to test Rosen's model to explain the pay gap for someone who was just promoted to a CEO position. In this thesis I've mainly investigated if the tournament model, developed by Rosen (1986) is applied by tournament directors or officials in the ATP World Tour. In order to see if the model is a reality I focused on three main issues: (1) are marginal payoffs and prize money levels structured according to tournament theory? (2) should the set format of a match influence the number of upsets seen? (3) how do spread differences and ranking levels affect the outcome of a match? In addition to these questions I observed how winning probabilities changes after the first set.

The results showed that marginal payoffs are not efficiently developed as Rosen's model predicts. Marginal payoff increases at a decreasing rate every round, and it decreases, in percentage terms, in the last round, almost reaching zero at times. The distribution of prize per round increases at a constant rate per round, and it mildly increases, in percentage terms, in the last round. Both structures of pay contradict the model.

Changes in the amount of sets played do influence the number of upsets seen in tournaments. Results showed that there are more upsets on matches played in best of 3 sets, when compared to matches played in best of 5 sets. Even though marginal payoffs and prize levels are not efficiently structured, a change in spread will influence the outcome of a match. An increase in spread difference will increase the number of upsets.

I cannot, on the other side, provide an answer that will tell the reader by how much effort level changes when there is an increase in spread. It is true to say that as differences in ranking increase, less competition is seen. As ranking differences decrease the number of upsets increase, and vice-versa happens. Lastly, it is also true to say that there is dominance among better players ranked in the top 10 compared to better players in other ranking subgroups.

This thesis was just a contribution that might help tournament directors reorganize prize money distribution, in order to get an optimal result in terms of profits and match quality. Better structure in spread could increase players' performance without increasing total purse, which might attract more audience. Perhaps, an interesting area for research would be to compare total purse levels with income levels: is tournament purse more important for better or worse ranked players?

APPENDIX

Table I. Means, standard deviations, minimum, and maximum values for regression variables. (n= 5,279)

Variable	Mean	Std. Dev.	Minimum	Maximum
Tournament	77.19	40.36	1	129
Year	2008.45	0.53	2007	2009
Sets	0.19	0.39	0	1
Round	1.93	1.16	1	7
Draw	57.65	37.19	28	128
Match	19.16	17.16	1	64
Top Ranking	38.59	36.75	1	557
Bottom Ranking	116.66	167.52	2	2000
First Set Winner	0.83	0.46	0	1
Outcome	0.67	0.46	0	1
Surface	1.55	0.71	1	4
Rank Difference	78.05	157.04	0	1999
Rank Weight	7874.59	4903.08	1869	18797
Average Ranking Weight	149.41	60.02	45.25	361.42
Prize Money Breakdown	21956.25	41122.44	1000	800000
Points	38.01	75.86	0	1200
Marginal Payoff	61789.85	81530.04	8836.17	850000
Total Prize Money Singles	2058177	2429653	272280	7547580
Winners Prize Money	443060.7	475142.1	64000	1600000
Maximum Rounds	5.642167	0.78	5	7

Final	0.024	0.15	0	1
Semi Final	0.04	0.21	0	1
Quarter Final	0.09	0.29	0	1
Top 16	0.19	0.39	0	1
Top 32	0.38	0.48	0	1
Top 64	0.15	0.36	0	1
Top 128	0.09	0.29	0	1
Hard Court	0.57	0.49	0	1
Clay Court	0.30	0.46	0	1
Grass Court	0.11	0.31	0	1
Carpet Court	0.005	0.07	0	1

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table II. Means, standard deviations, minimum, maximum, and percent change for variable MARGINAL PAYOFF, where i(round)= 1, 2,…, 5; DRAW = 28

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Marginal Payoff 1	15300.9	3842.78	10717.5	25073.75	34.01
Marginal Payoff 2	23186.94	6173.50	15865	39487.5	32.31
Marginal Payoff 3	34258.25	9379.65	23500	60075	23.48
Marginal Payoff 4	44774.47	12816.59	30000	79650	-2.17
Marginal Payoff 5	43822.78	12922.16	30000	78300	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table III. Means, standard deviations, minimum, maximum, and percent change for variable MARGINAL PAYOFFi, where i(round)= 1, 2,…, 5; DRAW = 32

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Marginal Payoff 1	21438	13367.64	11177.5	71300	36.01
Marginal Payoff 2	33502.48	21838.52	17406.25	118200	33.80
Marginal Payoff 3	50613.25	34581.17	25412.5	185200	25.78
Marginal Payoff 4	68200.81	50898.97	32325	255000	0.22
Marginal Payoff 5	68354.37	53293.57	31450	270000	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table IV. Means, standard deviations, minimum, maximum, and percent change for variable MARGINAL PAYOFFi, where i(round)= 1, 2,…, 6; DRAW = 48

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Marginal Payoff 1	25407.04	11202.68	10443.75	47466	48.83
Marginal Payoff 2	49658.89	23981.07	15987.5	83160	36.17
Marginal Payoff 3	77799.36	37220.29	23775	121068	33
Marginal Payoff 4	116119.3	57966.79	33750	182655	30.61
Marginal Payoff 5	167361.4	82631.06	42000	245430	8.19
Marginal Payoff 6	182305	76212.16	42500	245430	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table V. Means, standard deviations, minimum, maximum, and percent change for variable MARGINAL PAYOFFi, where i(round)= 1, 2,…, 5; DRAW = 56

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Marginal Payoff 1	38201.96	17123.73	8836.172	69153.75	45.48

Marginal Payoff 2	70082.11	27244.06	19954.69	115762.5	36.25
Marginal Payoff 3	109937.1	43511.65	29851.88	187650	33.61
Marginal Payoff 4	165613.5	67821.27	41343.75	288225	25.72
Marginal Payoff 5	222985.6	95848.47	51198.75	399600	2.46
Marginal Payoff 6	228626.8	105682	46912.5	425250	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table VI. Means, standard deviations, minimum, maximum, and percent change for variable MARGINAL PAYOFFi, where i(round)= 1, 2,…, 5; DRAW = 96

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Marginal Payoff 1	63337.5	13262.87	55325	86125	38.78
Marginal Payoff 2	103465	20248.85	91010	138250	35.87
Marginal Payoff 3	161345	25326.75	145680	204500	31.82
Marginal Payoff 4	236650	24377.53	221500	277500	19.59
Marginal Payoff 5	294325	10875.23	277500	302400	3.5
Marginal Payoff 6	305000	8660.25	295000	310000	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table VII. Means, standard deviations, minimum, maximum, and percent change for variable MARGINAL PAYOFFi, where i(round)= 1, 2,…, 5; DRAW = 128

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Marginal Payoff 1	35021.93	3405.09	29942.69	39250	70.27
Marginal Payoff 2	117815.5	11289.03	100982.8	133000	40.92
Marginal Payoff 3	199443.4	21138	168155.9	232000	39.65
Marginal Payoff 4	330496.5	37310.14	281822.2	390000	34.03

Marginal Payoff 5	501041.3	60939.01	422730	600000	26.26
Marginal Payoff 6	679492.5	101641.1	563640	850000	-2.75
Marginal Payoff 7	661305	85780.53	563640	800000	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table VIII. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i(prize)= 1, 2,…, 5; DRAW = 28

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Prize 1	5094.95	1078.97	3600	8750	42.11
Prize 2	8802.38	1864.032	6385	14760	40.76
Prize 3	14860.21	3340.693	10500	25150	44.40
Prize 4	26731.22	6302.35	19000	43750	46.1
Prize 5	49594.31	12868.1	34000	81000	46.91
Prize 6	93417.08	25730.98	64000	159300	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table IX. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i(prize)= 1, 2,…, 5; DRAW = 32

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Prize 1	6094.783	2396	1000	14500	43.47
Prize 2	10781.54	4506.25	6350	26700	43.18
Prize 3	18977.39	9061.16	10600	52300	46.52
Prize 4	35490.24	19621.05	5550	110000	48.94
Prize 5	69513.87	42227.2	37000	230000	49.57
Prize 6	137868.2	95031.14	68800	500000	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table X. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i(prize)= 1, 2,…, 6; DRAW = 48

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Prize 1	7311.95	4678.05	3550	20250	34.08
Prize 2	11093.21	5005.56	6000	20452.5	49.23
Prize 3	21852.43	9434.10	10100	33034.5	47.46
Prize 4	41592.14	18177.77	17000	62775	43.81
Prize 5	74030.71	45179.32	5550	122715	76.83
Prize 6	319522.9	159940	94000	490860	9.66
Prize 7	353693.3	152849.5	94000	490860	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XI. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i(prize)= 1, 2,…, 6; DRAW = 56

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Prize 1	10836.18	5266.68	3915	20250	26.14
Prize 2	14672.08	5465.57	7425	23625	50.74
Prize 3	29785.67	9758.65	12825	45562.5	47.66
Prize 4	56916	19589.45	22005	89100	48.74
Prize 5	111036.7	40205.86	37395	177525	49.46
Prize 6	219708.8	85769.72	64800	364500	50.99
Prize 7	448335.7	191104.7	113805	789750	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XII. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i(prize)= 1, 2,…, 6; DRAW = 96

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Prize 1	13395	3862.14	10580	20000	46.42
Prize 2	25000	6992.78	20400	37000	47.69
Prize 3	47792.5	14795.23	38570	73000	47.37
Prize 4	90812.5	28596.84	73500	138750	49.63
Prize 5	180300	59993.71	147500	277500	38.95
Prize 6	295333.3	288.67	295000	295500	50.8
Prize 7	600333.3	8948.92	590000	605500	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XIII. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i(prize)= 1, 2,…, 7; DRAW = 128

Variable	Mean	Std. Dev.	Minimum	Maximum	Percentage Change
Prize 1	17894.09	1489.86	15682.5	20250	38.36
Prize 2	29030.21	2658.78	26010	33075	38.39
Prize 3	47123.97	4279.53	43031.25	54810	42.05
Prize 4	81319.1	6942.95	70452.8	92340	49.58
Prize 5	161295	14617.34	140910	178875	50
Prize 6	322590	29717.93	281820	357750	51.95
Prize 7	671430	86937.21	563640	800000	49.62
Prize 8	1332735	170330.9	1127280	1600000	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XIV. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i(prize)= 1, 2,…, 5; DRAW = 48 and YEAR = 2008

Variable	Mean	Std. Dev.	Min	Max	Percentage Change
Prize 1	5650	352.76	5300	6000	48.63
Prize 2	11000	0	11000	11000	41.25
Prize 3	18725	1363.03	17450	20000	43.25
Prize 4	33000	4618.80	29000	37000	45.45
Prize 5	60500	12020.82	52000	69000	47.16
Prize 6	114500	28991.38	94000	135000	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XV. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i= 1, 2,…, 6; DRAW = 48 and YEAR = 2008

Variable	Mean	Std. Dev.	Min	Max	Percentage Change
Prize 1	5181.25	1977.6	3550	8500	43.6
Prize 2	9187.5	4185.89	6000	16250	72.03
Prize 3	32852.25	188.22	32670	33034.5	47.38
Prize 4	62437.5	360.80	62100	62775	48.86
Prize 5	122107.5	701.48	121500	122715	50
Prize 6	244215	1718.26	243000	245430	50
Prize7	488430	3436.53	486000	490860	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XVI. Means, standard deviations, minimum, maximum, and percent change for variable PRIZE MONEY BREAKDOWNi, where i= 1, 2,…, 6; DRAW = 48 and YEAR = 2009

Variable	Mean	Std. Dev.	Min	Max	Percentage Change
Prize 1	5181.25	1977.6	3550	8500	43.6
Prize 2	9187.5	4185.89	6000	16250	46.37
Prize 3	17134.38	8337.79	10100	31000	48.57
Prize 4	33317.5	16308.87	17000	59200	37.57
Prize 5	53375	43694.24	5550	113725	60.45
Prize 6	134975	69884.21	51500	222000	52.85
Prize7	286325	144153.9	94000	443500	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XVII. Means, TOTAL PRIZE MONEY SINGLES, and % OF TOTAL PRIZE MONEY SINGLES, where DRAW = 28, 32, 48, 56, 96, and 128.

Variable	Mean	Variable	Mean
Winner Prize 28	93417.08	Winner Prize 32	137868.2
Finalist Prize 28	49594.31	Finalist Prize 32	69513.87
Semi Prize 28	26731.22	Semi Prize 32	35490.24
Total Purse 28	387473.3	Total Purse 32	538041
% Total Purse	**43.8**	**% Total Purse**	**45.140112**
Variable	Mean	Variable	Mean
Winner Prize 48	353693.3	Winner Prize 56	448335.7
Finalist Prize 48	319522.9	Finalist Prize 56	219708.8
Semi Prize 48	74030.71	Semi Prize 56	111036.7

Total Purse 48	1485727	Total Purse 56	1850889
% Total Purse	**50.29**	**% Total Purse**	**42.09**
Variable	Mean	Variable	Mean
Winner Prize 96	600333.3	Winner Prize 128	1332735
Finalist Prize 96	295333.3	Finalist Prize 128	671430
Semi Prize 96	180300	Semi Prize 128	322590
Total Purse 96	2321013	Total Purse 128	6773250
% Total Purse	**46.35**	**% Total Purse**	**34.35**
	Avg. % Total Purse	**43.67**	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XVIII. Percentages for FIRST SET WINNERi, and OUTCOMEi, where i=0 or 1.

First Set Winner	Outcome	Percentage
1	0	30.62
1	1	69.38
0	0	39.32
0	1	60.68

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XIX. Percentages for FIRST SET WINNERi where i=0 or 1.

First Set Winner	Percentage
0	19.09
1	79.64

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XX. Percentages for OUTCOMEi where i=0 or 1.

Outcome	Percentage
0	32.59
1	67.41

Source: Author’s calculations from data gathered on the ATP Tour website, 2010.

Table XXI. Marginal effects after logit regression on SETS and MARGINAL PAYOFF for the 2007-09 ATP Tour. Dependent variable= OUTCOME. Absolute z values in parenthesis.

Variable	(1)
Top Ranking	-.0023286 (9.97)
Bottom Ranking	.0004775 (6.56)
Clay Court	.0115025 (0.80)
Grass Court	.0153684 (0.71)
Carpet Court	-.0898073 (1.01)
Marginal Payoff 10k	**-.0020323 (2.06)**
Sets	**.0827337 (2.67)**
Top Ranking Grand Slam	-.0024583 (3.89)
Bottom Ranking Grand Slam	.0008218 (2.54)
(*) dy/dx is for discrete change of	dummy variable

Source: Author’s calculations from data gathered on the ATP Tour website, 2010.

Table XXII. Effect of total purse on entry into tournaments for the 2007-09 ATP Tour. Dependent variable= AVERAGE RANK WEIGHT. Absolute t-statistics in parenthesis.

Variable	(1)
Total Prize Money 10K	**-.142503 (4.37)**
Adj. R^2	0.1306
F value	19.07

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XXIII. Marginal effects after logit regression on RANKING DIFFERENCE for the 2007-09 ATP Tour. Dependent variable= OUTCOME. Absolute z values in parenthesis.

Variable	(1)
Clay Court	.0047745 (0.33)
Grass Court	.0027262 (0.13)
Carpet Court	-.0654743 (0.75)
Marginal Payoff 10k	.00154 (1.67)
Sets	.0638548 (3.64)
Ranking Difference	**.0004622 (6.32)**
(*) dy/dx is for discrete change of dummy variable from 0 to 1	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XXIV. Marginal effects after logit regression on PLAYERS' RANKING SUBGROUP for the 2007-09 ATP Tour. Dependent variable= OUTCOME Absolute z values in parenthesis.

Variable	(1)

Clay Court	.0120572 (0.83)
Grass Court	.0142309 (0.66)
Carpet Court	-.0683389 (0.77)
Marginal Payoff 10k	-.0026491 (2.26)
Sets	.0563614 (3.03)
Top 10 versus Bottom 11 to 25	**.0815007 (1.75)**
Top 10 versus Bottom 26 to 50	**.1089054 (2.45)**
Top 10 versus Bottom 51 to 75	**.1580663 (3.83)**
Top 10 versus Bottom 76 to 101	**.2113679 (5.65)**
Top 10 versus Bottom > 100	**.2092679 (5.80)**
Top 11 to 25 versus Bottom 11 to 25	**-.0946906 (1.30)**
Top 11 to 25 versus Bottom 26 to 50	**-.0484663 (0.83)**
Top 11 to 25 versus Bottom 51 to 75	**.009893 (0.17)**
Top 11 to 25 versus Bottom 75 to 101	**.0948124 (1.83)**
Top 11 to 25 versus Bottom > 100	**.1207259 (2.60)**
Top 26 to 50 versus Bottom 26 to 50	**-.0782457 (1.22)**
Top 26 to 50 versus Bottom 51 to 75	**-.1055609 (1.71)**
Top 26 to 50 versus Bottom 76 to 101	**-.0160428 (0.27)**
Top 26 to 50 versus Bottom > 100	**.0019647 (0.03)**
Top 51 to 75 versus Bottom 51 to 75	**-.1664179 (2.36)**
Top 51 to 75 versus Bottom 76 to 101	**-.1636106 (2.42)**

Top 51 to 75 versus Bottom > 100	**-.0745935 (1.19)**
Top 76 to 101 versus Bottom 76 to 101	**-.1351779 (1.81)**
Top 76 to 101 versus Bottom > 100	**-.106418 (1.65)**
Top > 100 versus Bottom > 100	**-.0977171 (1.49)**
(*) dy/dx is for discrete change of dummy variable 1	

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XXV. Logit regression on SETS and MARGINAL PAYOFF for the 2007-09 ATP Tour. Dependent variable= OUTCOME. Absolute z values in parenthesis.

Variable	(1)
Top Ranking	-.0107724 (9.93)
Bottom Ranking	.0022088 (6.50)
Clay Court	.0534163 (0.79)
Grass Court	.0718378 (0.71)
Carpet Court	-.3911111 (1.06)
Marginal Payoff 10k	**-.0094021 (2.06)**
Sets	**.402366 (2.52)**
Top Ranking Grand Slam	-.0113726 (3.88)
Bottom Ranking Grand Slam	. 0038017 (2.53)
_cons |	.8835156 (13.41)

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XXVI. Logit regression on RANKING DIFFERENCE for the 2007-09 ATP Tour. Dependent variable= OUTCOME. Absolute z values in parenthesis.

Variable	(1)
Clay Court	.0219219 (0.33)
Grass Court	.01252 (0.13)
Carpet Court	-.2869899 (0.78)
Marginal Payoff 10k	.00706 (1.66)
Sets	.3034473 (3.50)
Ranking Difference	**.0021189 (6.26)**
_cons		.4750604 (9.13)

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Table XXVII. Logit regression on PLAYERS' RANKING SUBGROUP for the 2007-09 ATP Tour. Dependent variable= OUTCOME Absolute z values in parenthesis.

Variable	(1)
Clay Court	.0561957 (0.83)
Grass Court	.0667021 (0.66)
Carpet Court	-.3019814 (0.80)
Marginal Payoff 10k	-.0122967 (2.26)
Sets	.2703712 (2.92)
Top 10 versus Bottom 11 to 25	**.4092022 (1.60)**
Top 10 versus Bottom 26 to 50	**.5627051 (2.15)**

Top 10 versus Bottom 51 to 75	**.8869497 (2.99)**
Top 10 versus Bottom 76 to 101	**1.344643 (3.60)**
Top 10 versus Bottom > 100	**1.313771 (3.78)**
Top 11 to 25 versus Bottom 11 to 25	**-.4128738 (1.36)**
Top 11 to 25 versus Bottom 26 to 50	**-.2178551 (0.85)**
Top 11 to 25 versus Bottom 51 to 75	**.0462823 (0.17)**
Top 11 to 25 versus Bottom 75 to 101	**.4840822 (1.63)**
Top 11 to 25 versus Bottom > 100	**.6350616 (2.23)**
Top 26 to 50 versus Bottom 26 to 50	**-.3450121 (1.27)**
Top 26 to 50 versus Bottom 51 to 75	**-.4613496 (1.79)**
Top 26 to 50 versus Bottom 76 to 101	**-.0735936 (0.27)**
Top 26 to 50 versus Bottom > 100	**.009133 (0.03)**
Top 51 to 75 versus Bottom 51 to 75	**-.706405 (2.48)**
Top 51 to 75 versus Bottom 76 to 101	**-.6966676 (2.55)**
Top 51 to 75 versus Bottom > 100	**-.330299 (1.24)**
Top 76 to 101 versus Bottom 76 to 101	**-.579212 (1.91)**
Top 76 to 101 versus Bottom > 100	**-.463918 (1.73)**
Top > 100 versus Bottom > 100	**-.4271677 (1.56)**
_cons |	**.7952297 (3.28)**

Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure IX: Percentage Average of Total Purse Distributed to Winners

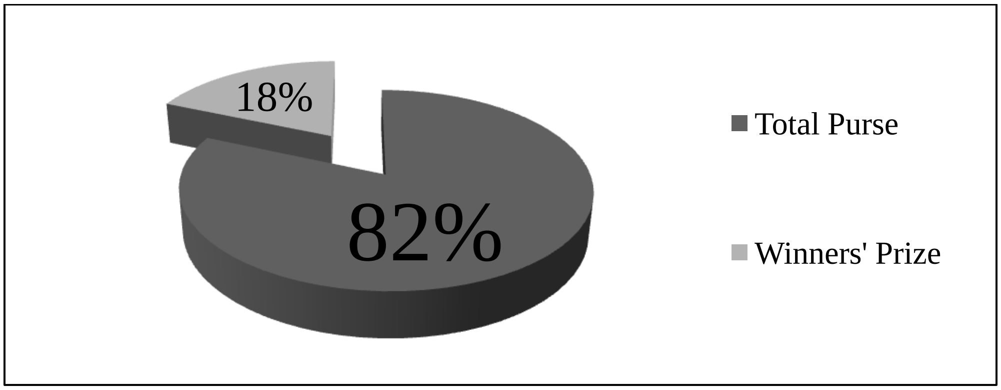


Figure X: Pure Strategy Equilibra

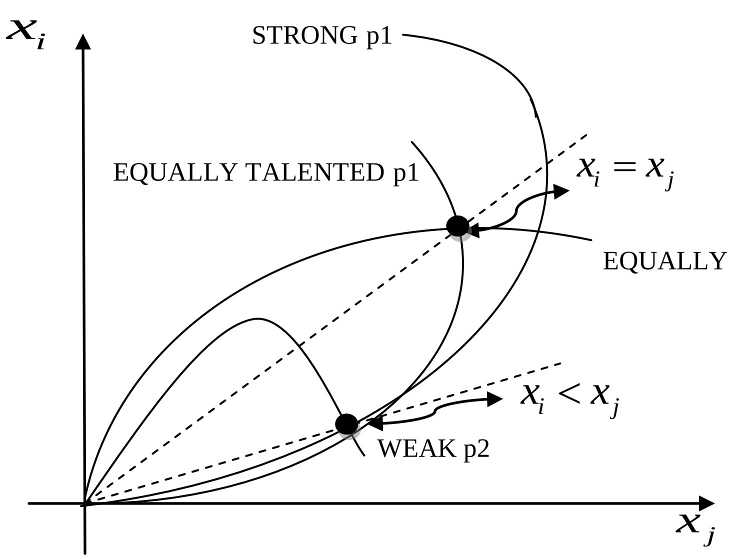
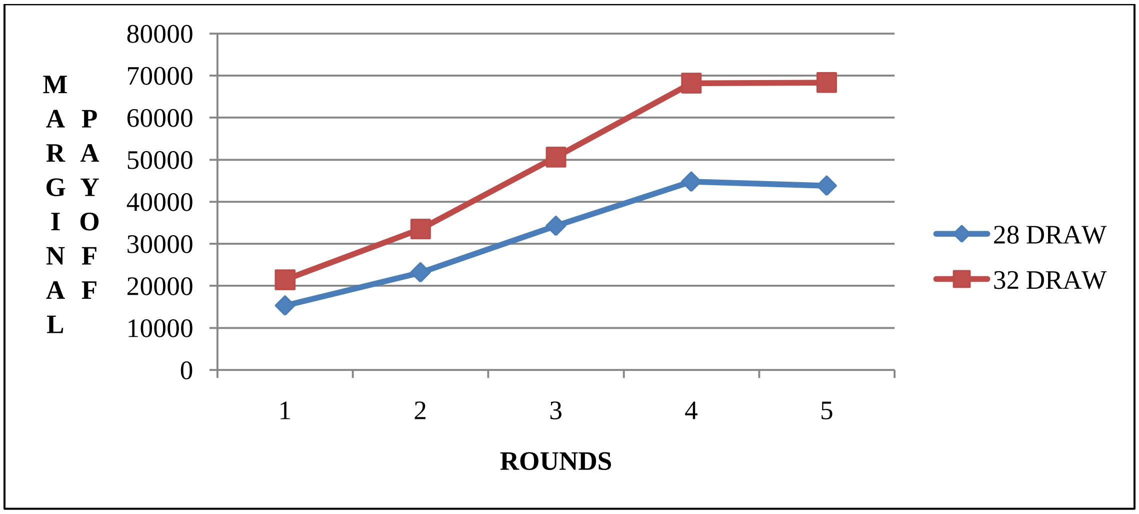
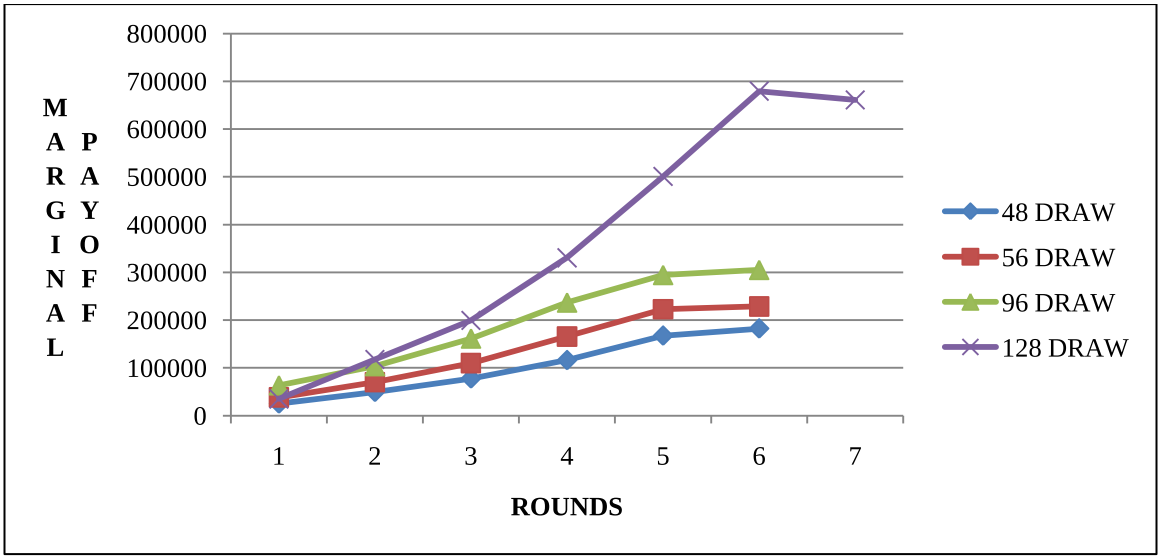


Figure XI: Mean Leves for MARGINAL PAYOFF when i=ROUND and DRAW= 28 and 32



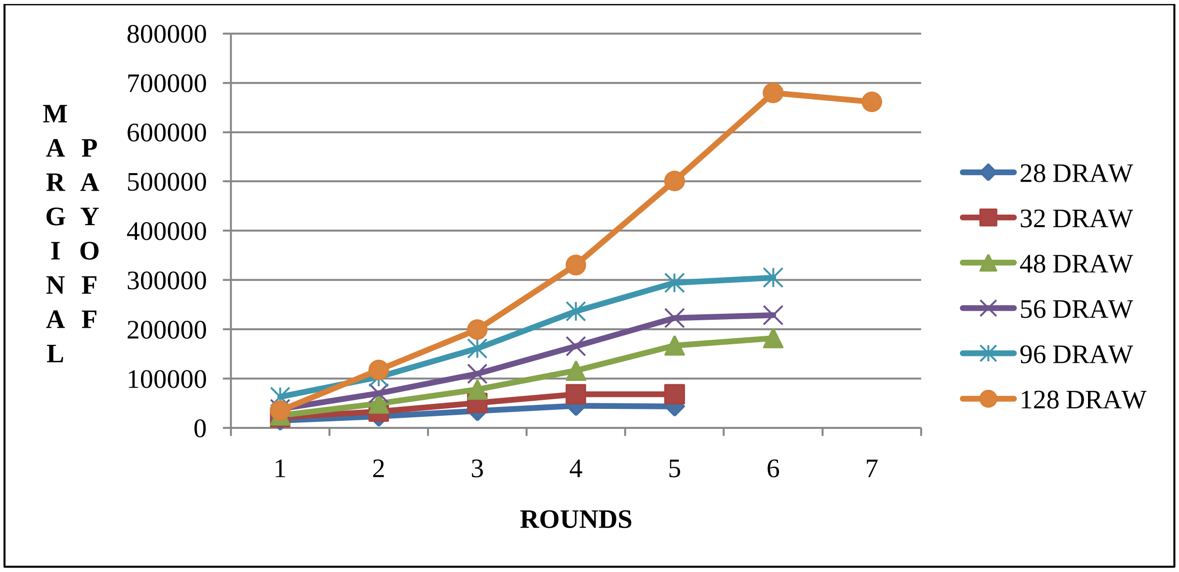
Source: Author’s calculations from data gathered on the ATP Tour website, 2010.

Figure XII: Mean levels for MARGINAL PAYOFF when i=ROUND and DRAW=49, 56, 96, AND 128



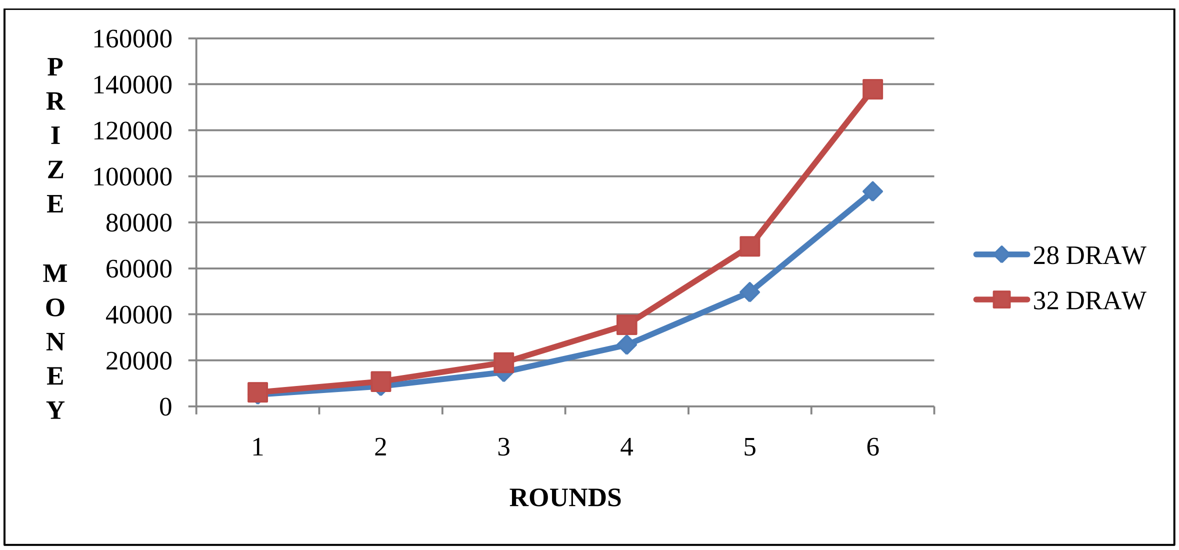
Source: Author’s calculations from data gathered on the ATP Tour website, 2010.

Figure XIII: Mean Leves for MARGINAL PAYOFF when i=ROUND and DRAW=28, 32, 48, 56, 96, and 128



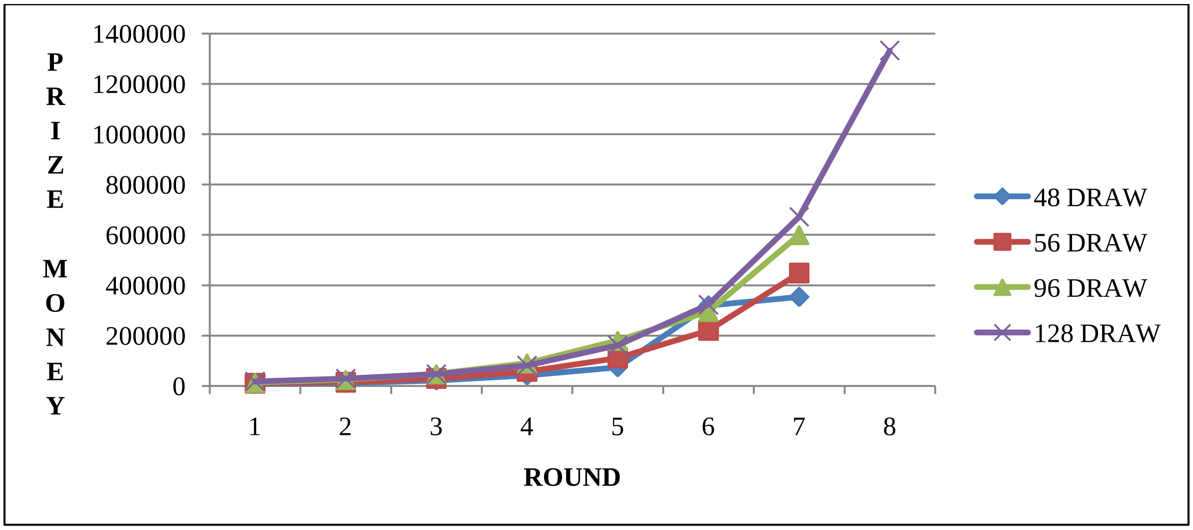
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure XIV: Mean Levels for PRIZE MONEY BREAKDOWN when i=PRIZE and DRAW=28 and 32



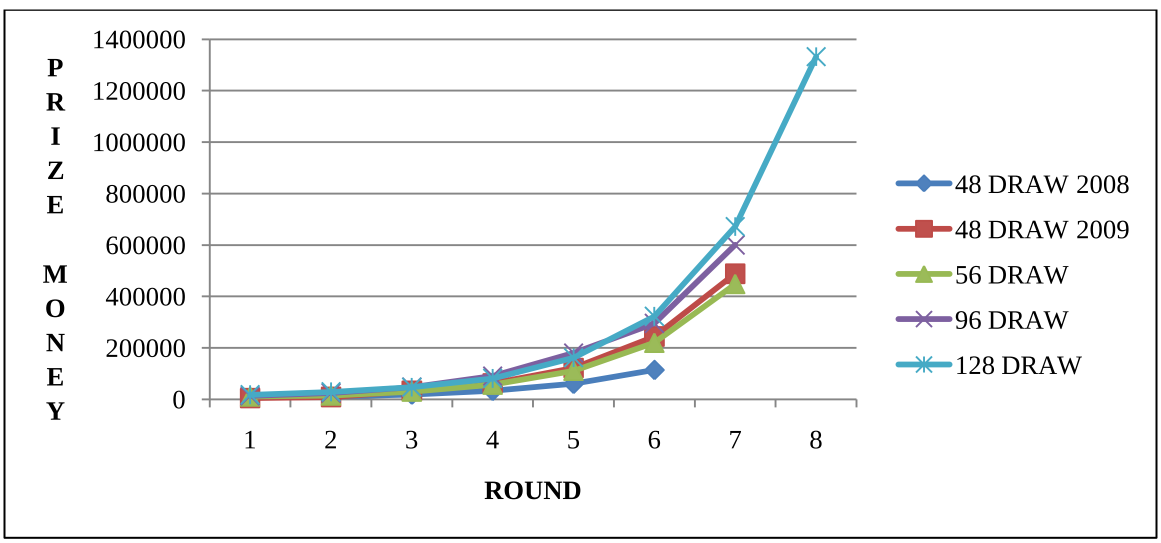
Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure XV: Mean Levels for PRIZE MONEY BREAKDOWN when i= PRIZE and DRAW= 48, 56, 96, and 128



Source: Author's calculations from data gathered on the ATP Tour website, 2010.

Figure XVI: Mean Levels for PRIZE MONEY BREAKDOWN when i=ROUND, DRAW=48 and YEAR=2008, DRAW=48 and YEAR=2009, and DRAW=56, 96, and 128



Source: Author's calculations from data gathered on the ATP Tour website, 2010.

BIBLIOGRAPHY

(2010) “The ATP World Tour,”ATP Tour, Inc. Web.

DeVaro, J., (2005) “Internal Promotion Competitions in Firms,” Econ WPA: Labor and Demography, no.0508005, pp.1-31

Dinterman, R., (2009) “Makings of a Tennis Upset,” Working Paper, Clemson University.

Eriksson, T., (1999) “Executive Compensation and Tournament Theory: Empirical Tests on Danish Data,” Journal of Labor Economics, Vol. 17, no. 2, pp. 262-280.

Green, J., Stokey, N., (1983) “A Comparison of Tournaments and Contracts,” Journal of Political Economy, Vol 91. No. 3, pp. 349-364

Ivankovic, M., (1995) “The Tournament Model: An empirical Investigation of the ATP Tour” PhD Thesis, Clemson University.

Knoeber, C., Thurman, W., (1994) “Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production,” Journal of Labor Economy, Vol 12. No. 2, pp.155-179

Lazear, P. E., (1996) “Performance Pay and Productivity,” NBER #5672.

Lazear, P. E., Rosen, S., (1981) “Rank-Order Tournaments as Optimum Labor Contracts,” Journal of Political Economy, Vol. 89, no. 5, pp. 841-864.

Lynch, J., (2000) “The Rewards to Running: Prize Structure and Performance in Professional Road Racing,” Journal of Sports Economics, no. 1, pp. 323-340.

Maloney, M. T., Bentley, C., (2010) “The Thrill of Victory: Measuring the Incentive to Win,” Journal of Labor Economics, Vol. 28, no.1, pp. 87-112.

Maloney, M.T., Terkun, K., (2002) “Road Warrior Booty: Prize Structures in Motorcycle Racing,” Contributions to Economic Analysis & Policy, Vol.1: Iss.1, Article 3. Available at: http://www.bepress.com/bejeap/contributions/vol1/iss1/art3

Main, B. G. M., O’Reilly III, C. A., Wade, J. (1993) “Top Executive Pay: Tournament or Teamwork?” Journal of Labor Economy, Vol. 11, no. 4, pp. 781-807.

Melton, M., Zorn, T., (2000) “An Empirical Test of Tournament Theory: The Senior PGA Tour,” Barmarick Publications (UK), Iss. 0307-4358.

Nalebuff, B. J., Stiglitz, J., (1983) "Prizes and Incentives: Toward a General Theory of Compensation and Compensation," Bell Journal of Economics, Vol 2, no.1, pp. 21-43

Rosen, S., (1986) "Prizes and Incentives in Elimination Tournaments," The American Economic Review, Vol. 76, no. 4, September, pp. 701-715.

Sunde, U., (2003) "Potential, Prizes and Performance: Testing Tournament Theory with Professional Tennis Data," Institute for the Study of Labor, no.947, pp. 2-28